People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

Echahid Cheikh Larbi Tebessi University – Tebessa

Faculty of Exact Sciences and Natural and Life Sciences

Department of Computer Science

Master Thesis

Submitted to the Department of the Computer Science in Partial Fulfillment of the Requirements for the degree Of Masters in Information Systems (IS)

Title

Embedded System for Corrosion Analysis and Prediction

Presented By

Raid BRAHMI

Jury Members:

Pr. Abdejalil GATTAL	Echahid Cheikh Larbi Tebessi University	**President**
Dr. Mounir ACHOURI	Echahid Cheikh Larbi Tebessi University	**Examiner**
Dr. Mohamed Yassine HAOUAM	Echahid Cheikh Larbi Tebessi University	**Supervisor**
Dr. Issam BENDIB	Echahid Cheikh Larbi Tebessi University	**Co-Supervisor**

Academic year: 2024/2025

Thanks

I would like to express my deep gratitude to **Dr. Haouam Mohamed Yassine**, my supervisor, for his invaluable support, insightful advice, and availability throughout this final project. His expertise and encouragement greatly contributed to the completion of this thesis.

I would also like to sincerely thank my Co-Supervisor, **Dr. IssamBendib**, for his guidance, helpful feedback, and continuous support during the course of this work. His contribution was instrumental in refining the direction and depth of this project.

My appreciation extends to the entire Computer Science teaching staff for the high-quality education and support I have received throughout my academic journey. I am also grateful to my classmates for their collaboration, motivation, and the enriching discussions we shared.

I would like to express my sincere thanks and congratulations to the members of the jury for their valuable time, attention, and constructive evaluation of my work. Their expertise and thoughtful remarks are deeply appreciated.

Finally, I dedicate this thesis to my family and all the people who provided moral support and encouragement throughout my academic career.

Thank you all very much.

Dedication

First, praise be to God always for all blessings, especially the blessing of knowledge.

To my **dear parents**,

For their love, sacrifices, and unconditional support throughout my university career.

May this work be an expression of my deep gratitude.

To **my brother and sisters**,

For their constant encouragement, kindness, and reassuring presence.

To my **loyal friends**,

For their moral support, invaluable help, and the moments we shared during these years of study.

I humbly dedicate this work to you.

BRAHMI RAID

Abstract

Corrosion is a major challenge in the oil and gas industry, compromising the safety, reliability, and sustainability of transportation infrastructure. This project proposes the design of an embedded system dedicated to analyzing and predicting corrosion in metal pipelines. It is based on the exploitation of data from inspections carried out over several years, combined with a BiLSTM (Bidirectional Long Short-Term Memory) deep learning model to predict the maximum corrosion depth at a given horizon. The dataset undergoes various preparation steps: missing value processing, correlation analysis and feature matching between inspection years. The developed model demonstrated high predictive accuracy with a low mean absolute error. This approach enables effective predictive maintenance and contributes to failure prevention, thus providing an innovative and suitable solution for real-time corrosion monitoring in embedded industrial systems.

Keywords: corrosion, pipelines, embedded system, BiLSTM, deep learning, prediction, inspection, predictive maintenance

Résumé

La corrosion constitue un enjeu majeur dans l'industrie pétrolière et gazière, compromettant la sécurité, la fiabilité et la durabilité des infrastructures de transport. Ce projet propose la conception d'un système embarqué dédié à l'analyse et à la prédiction de la corrosion dans les pipelines métalliques. Il repose sur l'exploitation de données issues d'inspections réalisées sur plusieurs années, combinées à un modèle d'apprentissage profond de type BiLSTM (Bidirectional Long Short-Term Memory) pour prédire la profondeur maximale de corrosion à un horizon donné. Le jeu de données est soumis à diverses étapes de préparation : traitement des valeurs manquantes, analyse de corrélation et correspondance des caractéristiques entre les années d'inspection. Le modèle développé a démontré une grande précision prédictive avec une erreur absolue moyenne faible. Cette approche permet une maintenance prédictive efficace et contribue à la prévention des défaillances, apportant ainsi une solution innovante et adaptée au suivi en temps réel de la corrosion dans les systèmes industriels embarqués.

Mots-clés :corrosion, pipelines, système embarqué, BiLSTM, deeplearning, prédiction, inspection, maintenance prédictive.

ملخص

يعد التآكل مشكلة رئيسية في صناعة النفط والغاز، حيث يعرض سلامة وموثوقية واستدامة البنية التحتية للنقل للخطر. يقترح هذا المشروع تصميم نظام مدمج مخصص لتحليل وتوقع التآكل في خطوط الأنابيب المعدنية. ويعتمد هذا النموذج على استخدام البيانات من عمليات التفتيش التي أجريت على مدى عدة سنوات، جنبًا إلى جنب مع نموذج التعلم العميق من نوع BiLSTM (الذاكرة طويلة المدى ثنائية الاتجاه) للتنبؤ بأقصى عمق للتآكل عند أفق معين. تخضع مجموعة البيانات لخطوات تحضيرية مختلفة: معالجة القيمة المفقودة، وتحليل الارتباط و مطابقة الميزات بين سنوات التفتيش. وأظهر النموذج المتطور دقة تنبؤية عالية مع متوسط خطأ مطلق منخفض. يتيح هذا النهج الصيانة التنبؤية الفعالة ويساهم في منع الفشل، وبالتالي توفير حل مبتكر ومناسب لمراقبة التآكل في الوقت الحقيقي في الأنظمة الصناعية المدمجة.

الكلمات المفتاحية: التآكل، خطوط الأنابيب، النظام المضمن، BiLSTM، التعلم العميق، التنبؤ، التفتيش، الصيانة التنبؤية.

Table of Content

List of Tables .....

List of Figures .....

General Introduction **1**

Problematic **2**

Objective and Contributions **2**

Thesis Outline **3**

Chapter 1: Fundamentals and General Concepts of Corrosion **4**

1. Introduction **5**

2. Definition of Corrosion **5**

3. Different Forms of Corrosion **6**

3.1 Uniform Corrosion 7

3.2 Localized Corrosion 7

3.3 Galvanic Corrosion 8

3.4 Stress Corrosion Cracking (SCC) 8

3.5 Erosion Corrosion 9

3.6 Filiform Corrosion (FFC) 9

4. Corrosion Influencing Factors **9**

4.1 Temperature 10

4.2 CO2 11

4.3 PH Values 11

4.4 Oxygen and Salinity 11

4.5 Metal Cations 11

4.6 Bacteria 12

5. Consequences of Corrosion **12**

5.1 Economic Consequences 12

5.1.1 Direct Costs 12

5.1.2 Indirect Costs 12

5.2 Environmental Consequences 12

5.2.1 Soil and Water Pollution 12

5.2.2 Impact on Ecosystems 12

5.3 Security Consequences 13
5.3.1 Industrial Risks 13
5.3.2 Public Infrastructure Risks 13
6. Prevention and Protection Methods 13
6.1 Cathodic Protection 14
6.2 Protective Coatings 14
6.3 Corrosion Inhibitors 14
6.4 Regular Maintenance and Monitoring 14
7. Automated Ultrasonic Testing (AUT) as a Data Source for Corrosion Prediction 15
7.1 Introduction to AUT in Corrosion Monitoring 15
7.2 Role of AUT in Corrosion Prediction 15
7.3 Advantages of Using AUT Data for Prediction 15
8. Conclusion 16
Chapter 02: 17
Stat of the Art and Corrosion Prediction Models 17
1. Introduction 18
2. Historique of Prediction Methods in Oil and Gas Industry 19
3. Physical Models 20
4. Statistical Models 20
5. Machine Learning Models 20
5.1 Artificial Neural Networks (ANN) 22
5.2 Support Vector Machine (SVM) 23
5.3 Random Forest (RF) 23
5.4 Decision Tree 23
6. Deep Learning Models 24
6.1 Deep Neural Network (DNN) 25
6.2 Recurrent Neural Network (RNN) 25
6.3 Long Short-Term Memory (LSTM) 25
6.4 Temporal Convolutional Networks 26
7. Further Work and Research Gaps 27
7.1 Recent ML and DL Approaches for Corrosion Prediction in Oil and Gas Pipelines 27
7.2 Summary Table 29

7.3 Discussion 31
8. Conclusion 32
Chapter 03: 34
Implementation and Evaluation of the Prediction Model 34
1. Introduction 35
2. Dataset Description 35
3. Framework and Implementation 37
4. Data Cleaning 38
4.1 Handling Missing Values 38
4.2 Correlation Matrix 40
5. Spatial Matching and Feature Tracking Across Inspections 44
6. Implementation and Results 46
6.1 Bidirectional LSTM for Corrosion Depth Prediction 46
6.2 Data Preparation 46
6.3 Model Architecture 46
6.4 Model summary 47
6.5 Training Configuration 47
6.6 Model Performance 48
6.7 Discussion of Prediction Results 48
7. Embedded System Using Raspberry Pi 4 49
7.1 Embedding the BiLSTM Model for Edge Inference 49
7.2 Dataset Partitioning and External Storage 50
7.3 Optimized Spatial Matching and Feature Tracking per Segment 50
7.4 Benefits of Raspberry Pi-Based Deployment 50
8 Conclusion 51
General Conclusion 52
Bibliographies 55

List of Tables

Table 1.1: Pipeline Corrosion Prediction Workflow Based on AUT Data. 15

Table 2.1: Summary of Research Works Using ML/DL for Corrosion Prediction in the Oil and Gas Industry..... 30

List of Figures

Figure 1.1: A simple illustration of the various forms of corrosion. 6
Figure 1.2: Pitting corrosion. 7
Figure 1.3: Schematic representation of galvanic corrosion. 8
Figure 1.4: Intergranular and transgranular stress corrosion cracking of the AISI 316L stainless steel at polythionic acid 8
Figure 1.5: Schematic diagram of corrosion in oil and gas industry 10
Figure 1.6: Corrosion by temperature 10

Figure 2.1: First application time and brief development history of the three methods 19
Figure 2.2: Machine learning models applied for corroded oil and gas pipelines integrity. 21
Figure 2.3: Statistics of benchmark models in the field of oil and gas pipeline corrosion prediction. 22
Figure 2.4: The artificial neural network architecture. 22
Figure 2.5: Decision Tree Example. 24
Figure 2.6: Bidirectional LSTM architecture 26

Figure 3.1: Framework for Corrosion Prediction Using BiLSTM. 37
Figure 3.2: Year1 -Missing values-. 38
Figure 3.3: Year3 -Missing values. 39
Figure 3.4: Year5 -Missing values-. 39
Figure 3.5: Year7 -Missing values-. 40
Figure 3.6: Correlation Matrix for year 1. 41
Figure 3.7: Correlation Matrix for year 3. 42
Figure 3.8: Correlation Matrix for year 5. 42
Figure 3.9: Correlation Matrix for year 7. 43
Figure 3.10: Distribution of SIPRD.m and SIPO.deg for GWNUM 22. 45
Figure 3.11: GWNUM 22 - Matched corrosions. 46
Figure 3.12: BiLSTM model structured. 47
Figure 3.13: Predicted vs Actual Corrosion Depth (MD.mm) at Year 7. 49

General Introduction

Corrosion represents a critical challenge in the oil and gas industry, particularly affecting metal infrastructure such as pipelines. These pipelines, which play a central role in transporting oil and gas over long distances, are constantly exposed to harsh environmental and operational conditions including humidity, pressure variations, chemical agents, and temperature fluctuations. Over time, such exposure leads to progressive material degradation, known as corrosion, which compromises structural integrity and increases the risk of leaks, ruptures, environmental contamination, and even catastrophic industrial accidents. As the global energy demand continues to grow, the reliability and safety of pipeline networks have become a strategic priority for operators and governments alike.

Historically, pipeline corrosion has been addressed using physical models and periodic manual inspections. While these approaches provide useful assessments, they often suffer from limitations in scalability, timeliness, and accuracy especially in predicting future degradation. Physical and statistical models, although widely used, typically require extensive domain knowledge and often struggle to handle non-linear, high dimensional, and incomplete datasets generated from inspection campaigns. Furthermore, these models fail to capture the temporal dynamics of corrosion progression over time, making them less suitable for predictive maintenance.

With the rapid development of artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL) techniques, new data-driven approaches have emerged to address these limitations. Among them, the Bidirectional Long Short Term Memory (BiLSTM) model has gained significant attention for its ability to model complex temporal sequences and predict future outcomes with high precision. BiLSTM networks are especially suited to time-series data, such as those generated by periodic inspections, as they can learn from both past and future patterns simultaneously offering a more robust prediction framework.

In this context, the integration of deep learning models into embedded systems provides a powerful and scalable solution for real-time corrosion analysis and prediction. Embedded

systems offer the advantage of being lightweight, low power, and deployable directly on industrial sites. By embedding an intelligent BiLSTM based prediction model into such a system, it becomes possible to continuously monitor pipeline health, detect early signs of severe

corrosion, and anticipate future failures enabling predictive maintenance strategies that reduce costs and improve safety.

This project proposes the design and development of an embedded system dedicated to the analysis and prediction of corrosion in oil and gas pipelines. Based on historical inspection data, the system uses a BiLSTM deep learning model to predict the maximum corrosion depth several years into the future. The data pipeline involves rigorous preprocessing steps including missing value treatment, outlier detection, correlation analysis, and spatial feature matching across multiple inspection years. The system aims not only to enhance prediction accuracy but also to enable proactive decision-making, reduce operational risks, and support the long-term sustainability of critical energy infrastructure.

Problematic

How can we design a system capable of accurately predicting corrosion development in oil and gas pipelines, given the increasing complexity of inspection data (nonlinear, noisy, and temporal), while traditional monitoring methods based on physical or statistical models are reaching their limits?

Objective and Contributions

The main objective of this thesis is to develop an embedded system capable of analyzing and predicting corrosion in metallic oil and gas pipelines using deep learning techniques. To achieve this, the project makes the following scientific and practical contributions:

Contribution 1: Feature Matching Method

An automated spatial feature matching method was developed to track corrosion anomalies across multiple inspection years. This method improved spatial consistency and enabled accurate temporal alignment of pipeline features, laying the groundwork for effective time-series prediction.

Contribution 2:Corrosion Prediction Model

A predictive model based on Bidirectional Long Short-Term Memory (BiLSTM) networks was implemented to estimate the maximum corrosion depth at year 7. This model achieved high accuracy and supports predictive maintenance strategies, helping to reduce operational risk and improve the safety and reliability of pipeline infrastructure.

Thesis Outline

The remainder of this thesis is organized as follows:

- **Chapter 1** presents the fundamentals and general concepts of corrosion, including its types, causes, and impacts on oil and gas pipelines.
- **Chapter 2** reviews the state of the art in corrosion prediction models, including physical, statistical, and AI-based approaches.
- **Chapter 3** describes the implementation of the proposed BiLSTM prediction model, the data processing steps, and the performance evaluation.
- Finally, the **General Conclusion** summarizes the key findings and outlines future perspectives.

Chapter 1: Fundamentals and General Concepts of Corrosion

1. Introduction

Corrosion is a natural phenomenon characterized by the gradual deterioration of materials, particularly metals, due to their environment. This primarily electrochemical process poses a major challenge in many industrial sectors such as civil infrastructure, the oil industry, aeronautics, and maritime transport. Due to its considerable economic and environmental impact, corrosion represents a global problem requiring effective prevention and control strategies(Koch et al., 2016).

Scientifically, corrosion is based on redox reactions involving the oxidation of the metal (anodic reaction) and the reduction of agents present in the environment (cathodic reaction)(Li et al., 2019). These reactions are influenced by several factors, including the nature of the material, climatic conditions (humidity, temperature, salinity), and the mechanical stresses experienced by metal structures.

Corrosion can take various forms: uniform corrosion, affecting the entire surface of the metal, localized corrosion, manifesting as deep pits or crevices, or galvanic corrosion, which results from contact between two metals of different types(Raja & Shoji, 2021) Each type of corrosion presents specific challenges in terms of detection and prevention .

To address these issues, recent advances in monitoring technologies have enabled the emergence of innovative solutions such as smart embedded systems. These technologies, integrating advanced sensors and artificial intelligence models, facilitate the early detection of corrosion and optimize predictive maintenance strategies.

2. Definition of Corrosion

Corrosion is a natural process of degradation of materials, mainly metals, resulting from their interaction with their environment. This phenomenon is generally of chemical, electrochemical or physical origin, causing a progressive alteration of the mechanical and aesthetic properties of the material . Loss of infrastructure efficiency (Ravi Kumar Gali & Ramesh Kumar V, 2023).

Electrochemically, corrosion is often caused by oxidation and reduction reactions. These Reactions Typically Involve:

- An anodic zone where the metal loses electrons (oxidation).
- A cathode zone where reduction reactions occur, often in the presence of oxygen or hydrogen ions.

Corrosion can occur in a variety of environments, including humid, marine, industrial, and underground environments. Its effects can manifest as rust stains, pitting, cracks, or a gradual loss of material thickness(H. Wang, 2023).

3. Different Forms of Corrosion

Corrosion mechanisms refer to the processes of progressive degradation of materials, particularly metals, resulting from chemical, electrochemical or physical interactions with their environment, which can compromise structural integrity and require solutions such as specialized materials or protective coatings to limit these effects (Prasetia, 2025).

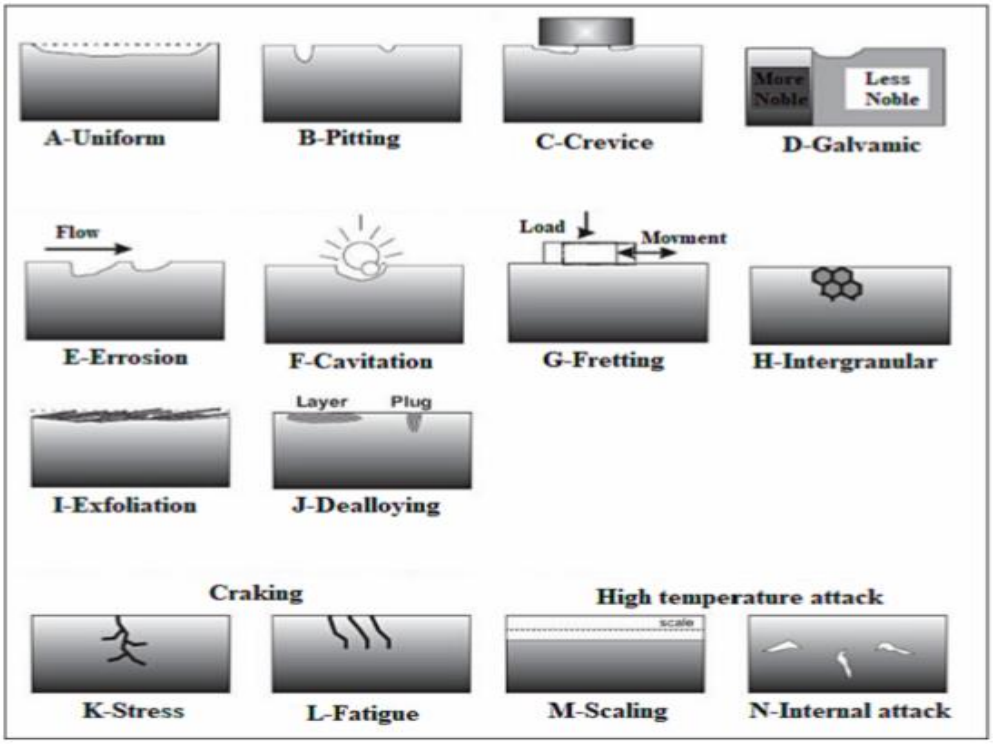


Figure 1.1: A simple illustration of the various forms of corrosion. (Bardal, 2004)

3.1 Uniform Corrosion

Uniform corrosion, also known as generalized corrosion, refers to a degradation process that affects the entire metal surface homogeneously, resulting from a balanced distribution of anodic and cathodic reactions. Although this form of corrosion leads to a progressive thinning of the material, its predictable nature allows its service life to be accurately estimated and managed effectively by integrating a corrosion margin into the design of metal structures (Khoshnaw & Gubner, 2020).

3.2 Localized Corrosion

Localized corrosion, unlike generalized corrosion, occurs in restricted areas of chemical processing equipment. For example, pitting corrosion in a chemical transfer pipeline is a good example of this phenomenon. This form of corrosion progresses rapidly through the metal, often at rates much higher than those of uniform corrosion. Such degradation can lead to leaks of flammable or explosive substances from equipment, storage tanks, or transfer lines, increasing the risk of fire or explosion, with serious consequences for human life and property (Tait, 2018) .



Figure 1.2: Pitting corrosion**.**

(Khoshnaw & Gubner, 2020b)

3.3 Galvanic Corrosion

Galvanic corrosion occurs when two dissimilar metals are in contact in the presence of an electrolyte, causing the more active metal (anode) to corrode faster while the more resistant metal (cathode) corrodes slower, with the corrosion rate increasing as the metals are further apart in the galvanic series (Hamdy Makhlouf, 2015).

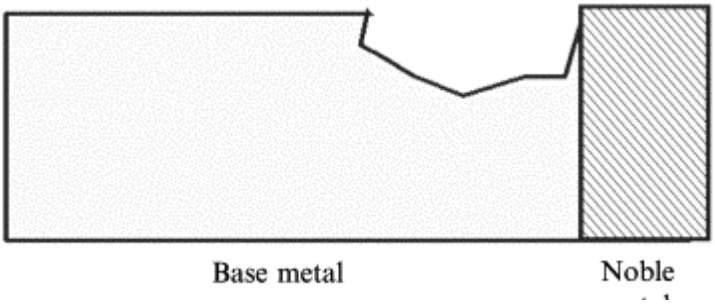


Figure 1.3: Schematic representation of galvanic corrosion.

(Hamdy Makhlouf, 2015)

3.4 Stress Corrosion Cracking (SCC)

Stress corrosion cracking (SCC) occurs when materials under tensile stress are exposed to specific corrosive environments, causing microscopic cracks that may not be visible but can grow rapidly and lead to sudden, catastrophic failures(Khalifeh, 2019).

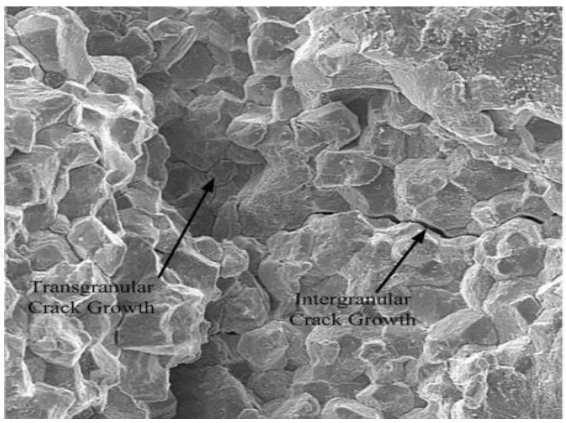


Figure 1.4: Intergranular and transgranular stress corrosion cracking of the AISI 316L stainless steel at polythionic acid (Khalifeh, 2019)

3.5 Erosion Corrosion

Erosion corrosion is a complex phenomenon that occurs when materials exposed to corrosive fluids experience simultaneous mechanical erosion and electrochemical corrosion, often resulting in greater material loss than either process alone. This interaction, known as the synergistic effect, occurs when corrosion accelerates erosion, while the additive effect describes how erosion enhances corrosion. Factors influencing this process include solid particle characteristics (such as size, shape, and velocity), the properties of the target material (such as hardness and toughness), and environmental conditions (such as slurry composition, flow velocity, and temperature) (Zadeh, 2023).

3.6 Filiform Corrosion (FFC)

Filiform corrosion (FFC) is a form of underpaint corrosion affecting metals like aluminum and magnesium, characterized by filament-like growth along the metal/coating interface. It involves an anodic region, a cathodic tail, and solid corrosion product accumulation. FFC appears as either rapid anodic undercutting or successive pitting. Studies on aluminum alloys reveal that surface composition and microstructure influence FFC behavior, with hydrogen evolution and bubble detachment near the anodic region playing a key role in coating failure. Effective prevention includes surface treatments such as anodizing, etching, desmutting, and conversion coating, particularly for aluminum containing post-consumer scrap (Mysliu et al., 2025).

4. Corrosion Influencing Factors

The deterioration of buried metal pipes due to corrosive soil conditions remains a significant global concern despite advancements in corrosion protection technologies. Understanding the key factors influencing this corrosion is crucial. These factors include moisture content, soil resistivity, pH, dissolved oxygen, temperature, and microbial activity. Additionally, the manufacturing method and material type, such as cast iron, ductile iron, and mild steel, impact corrosion behavior. Corrosion rates tend to increase with moisture content up to a critical level, while higher soil resistivity generally reduces corrosion risk. Organic-rich soils exhibit higher resistivity, and dissolved oxygen accelerates corrosion by forming differential cells. Furthermore, bacteria in soil create biofilms that progressively degrade metallic pipes over time (Wasim et al., 2018).

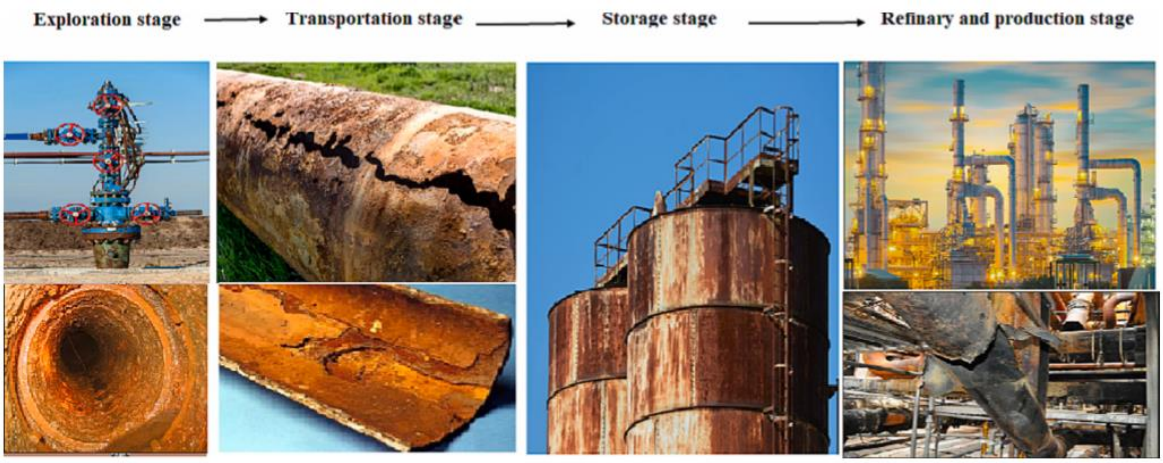


Figure 1.5: Schematic diagram of corrosion in oil and gas industry (Hussein Khalaf et al., 2024).

4.1 Temperature

Corrosion rates increase with temperature as diffusion processes accelerate. In ships, areas near engine rooms, heated cargo tanks, or under-deck spaces tend to corrode faster. In modern double-hulled tankers, empty ballast tanks act like a thermos, retaining heat longer and creating temperature differences that accelerate corrosion in ballast and cargo tanks (Factors Influencing Rate of Corrosion - Sealxpert, 2017).

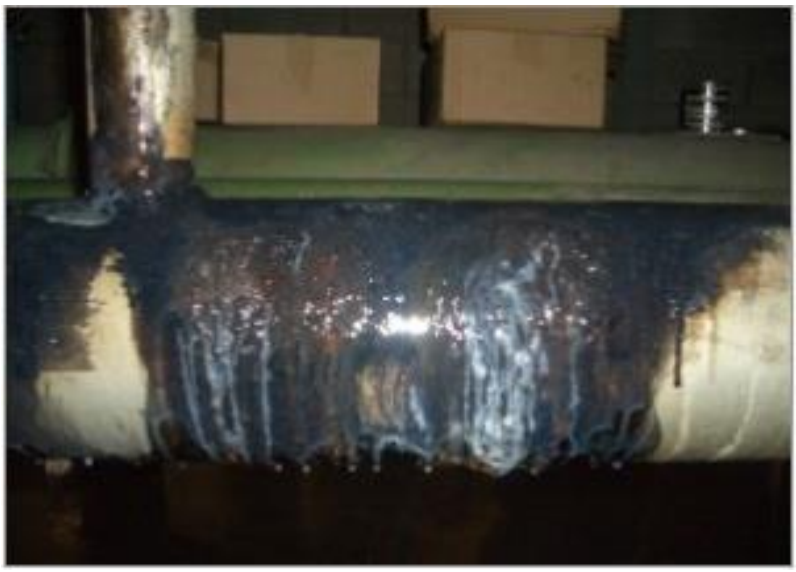


Figure 1.6: Corrosion by temperature (Factors Influencing Rate of Corrosion - Sealxpert, 2017)

.

4.2 CO2

CO_2 concentration and partial pressure are key factors in the onset of corrosion. According to the literature, when the partial pressure is below 0.02 MPa, corrosion is low and negligible; however, above 0.2 MPa, corrosion becomes significant. CO_2 strongly influences the corrosion rate, mainly by modifying the pH and the density of the corrosion product film. The higher the CO_2 concentration, the more acidic the fluid becomes, leading to a decrease in pH and an intensification of the corrosion phenomenon (Hussein Khalaf et al., 2024).

4.3 PH Values

Copper corrosion in aqueous solutions varies with pH, with the formation of more compact and effective protective oxide films at high pH (up to CuO at pH 12-13), while at low pH (about 3), porous corrosion products promote diffusion and accelerate copper degradation (Feng et al., 1997).

4.4 Oxygen and Salinity

Oxygen plays an important role in the corrosion mechanisms of metals and their alloys in aerated environments. In aerated seawater, the concentration of dissolved oxygen is highly variable since it depends on the following physical, chemical and biological factors: exchanges at the air/ocean interface, diffusions and mixing within the water mass, photo-oxidation phenomena, chemical oxidation reactions, etc. The dissolution of oxygen is governed by laws (Mémoire de Fin d études, n.d.).

Salinity, characterized by the presence of ions such as sodium (Na^+) and chloride (Cl^-), increases the electrical conductivity of water, enhancing the electrochemical reactions responsible for corrosion. For instance, marine environments rich in chloride ions are highly corrosive to metals like steel and aluminum (Ch10-La-Tenue-a-La-Corrosion-de-l-Aluminum-En-Milieu-Marin.Pdf, n.d.).

4.5 Metal Cations

Metal cations such as Na^+, K^+, and Mg^{2+} influence the corrosion rate of carbon steel, with Mg^{2+} accelerating corrosion more significantly than Na^+ and K^+. This is due to Mg^{2+}'s ability to enhance steel dissolution and promote the formation of β-FeOOH, which reduces the effectiveness of the protective layer. Additionally, the corrosion current density increases in the

presence of Mg^{2+}, while polarization resistance and charge transfer resistance decrease (Liu et al., 2020).

4.6 Bacteria

Among the microorganisms that induce biocorrosion, bacteria are the first to count upon. Due to their smaller size and higher surface-to-volume ratio, bacteria can carry out chemical reactions much more rapidly than larger organisms. Bacterial metabolic activities degrade organic matter into simpler products, which are then used to produce biomass. These metabolic processes change the immediate environment, which leads to the material corroding (Rao & Mulky, 2023).

5. Consequences of Corrosion

5.1 Economic Consequences

Corrosion has significant economic consequences, manifesting itself in direct and indirect costs (Léopold, 2024).

5.1.1 Direct Costs

- Replacement of damaged equipment.
- Maintenance and repair costs.

5.1.2 Indirect Costs

- Production losses due to maintenance shutdowns.
- Increased risk of industrial accidents.

5.2 Environmental Consequences

5.2.1 Soil and Water Pollution

The deterioration of metal infrastructure, such as pipelines and reservoirs, can release heavy metals and other toxic substances into the environment. These contaminants can leach into soil and groundwater, compromising drinking water quality and affecting human health (Corrosion, n.d.).

5.2.2 Impact on Ecosystems

The presence of corrosive substances and corrosion products in the environment can disrupt natural habitats, leading to a loss of biodiversity. Animal and plant species exposed to these

contaminants can suffer adverse effects, ranging from physiological disorders to mortality (Sylvain.S.39, 2024).

5.3 Security Consequences

Corrosion is a critical issue that significantly impacts both public and industrial safety. It can lead to severe hazards such as explosions, fires, and toxic leaks, while also adversely affecting public infrastructure. Economically, the global cost of corrosion is estimated to be approximately $2.5 trillion annually, equating to about 3.4% of the world's GDP(Zehra et al., 2022).

5.3.1 Industrial Risks

- **Equipment Failure:** Corrosion can compromise the structural integrity of industrial equipment, leading to leaks of hazardous or flammable materials. This increases the risk of explosions or fires within industrial facilities(Joseph et al., 2022).
- **Catastrophic Accidents:** In sectors like chemical and oil industries, the failure of corroded equipment can result in catastrophic incidents, endangering workers' lives and causing severe environmental and economic damage(Joseph et al., 2022).

5.3.2 Public Infrastructure Risks

- **Structural Failures:** Corrosion of bridges, columns, and buried pipelines can lead to sudden structural collapses, posing direct threats to public safety (W. Xu et al., 2024).
- **Service Disruptions:** Corrosion in water and gas distribution systems can cause contamination or interruption of essential services, disrupting daily life and leading to significant economic losses (Kashgain, n.d.).

6. Prevention and Protection Methods

Corrosion prevention and protection methods are crucial for maintaining the integrity and longevity of infrastructure across various industries. These methods aim to mitigate the electrochemical reactions that lead to material degradation. The primary strategies include cathodic protection, protective coatings, corrosion inhibitors, and regular maintenance. Each method has its own advantages and limitations, and often a combination of these strategies is employed to achieve optimal results.

6.1 Cathodic Protection

- Cathodic protection (CP) is a widely used method, especially in the oil and gas industry and marine environments, to prevent corrosion by making the metal surface the cathode of an electrochemical cell (Sheryazov et al., 2024).
- It is effective in environments where the metal is in contact with an electrolyte, such as soil or water, and is often used in combination with coatings for enhanced protection (Hossen et al., 2023)

6.2 Protective Coatings

- Protective coatings act as a physical barrier to prevent corrosive substances from reaching the metal surface. They are commonly used in both the oil and gas industry and marine environments (Hossen et al., 2023).
- In concrete structures, sealing the surface can minimize the ingress of harmful ions and gases, thereby protecting the embedded steel (Poursaee, 2023).

6.3 Corrosion Inhibitors

- Corrosion inhibitors are chemicals that, when added to the environment, decrease the corrosion rate. They are particularly useful in closed systems where they can be easily maintained (Sheryazov et al., 2024) .
- These inhibitors can be applied in various forms, including paints with specific chemical compositions (Sooriyalakshmi & Helena, 2021) .

6.4 Regular Maintenance and Monitoring

- Regular maintenance and monitoring are essential to detect early signs of corrosion and take corrective actions. Techniques such as linear polarization and impedance measurement are used for monitoring (Sooriyalakshmi & Helena, 2021) .
- Continuous improvement and innovation in monitoring technologies are necessary to enhance the effectiveness of corrosion prevention strategies (Sheryazov et al., 2024) .

While these methods are effective, they are not without challenges. The choice of method depends on factors such as environmental conditions, cost, and the specific requirements of the infrastructure. Moreover, the development of more environmentally friendly and sustainable corrosion prevention technologies remains a significant area of research (Zehra et al., 2022) .

7. Automated Ultrasonic Testing (AUT) as a Data Source for Corrosion Prediction

7.1 Introduction to AUT in Corrosion Monitoring

Automated Ultrasonic Testing (AUT) is a non destructive inspection technique widely used in the oil and gas industry to evaluate the integrity of pipelines. AUT systems generate precise and high-resolution measurements of critical parameters such as wall thickness, corrosion depth, feature length, and feature width. These measurements serve as essential inputs for corrosion monitoring, maintenance planning, and predictive modeling.

In this project, AUT data plays a central role not only in identifying the current state of pipeline health but also in building accurate predictive models using Machine Learning (ML) and Deep Learning (DL) algorithms (Z. Wang et al., 2022).

7.2 Role of AUT in Corrosion Prediction

The role of AUT within a corrosion prediction framework can be summarized in the following key steps:

Table 1.1: Pipeline Corrosion Prediction Workflow Based on AUT Data.

Step	Description
Data Acquisition	AUT systems collect high-frequency ultrasonic data across pipeline surfaces, detecting defects and wall thinning.
Feature Extraction	Important features are extracted from AUT datasets, such as maximum depth (MD.mm), feature width (W.mm), and feature length (L.mm).
Data Preprocessing	The extracted data is cleaned, missing values are handled, and outliers are detected and treated.
Model Training	ML/DL models (such as LSTM) are trained on historical AUT data to learn complex degradation patterns over time.
Prediction	Trained models predict future corrosion severity, maximum pitting depth, or potential burst risks in pipelines

7.3 Advantages of Using AUT Data for Prediction

- ✓ **High Accuracy:** AUT provides precise, reliable thickness and defect measurements.
- ✓ **Automation:** Reduces human error and enables large-scale inspections.
- ✓ **Data-Driven Insights:** Enables machine learning models to learn from real-world corrosion behavior.

- ✓ **Early Risk Detection:** Facilitates predictive maintenance before critical failures occur.

8. Conclusion

In this chapter, we explored the fundamental aspects of corrosion, including its mechanisms, influencing factors, and the significant consequences it poses on economic, environmental, and safety levels. The integration of advanced technologies, particularly machine learning algorithms, has shown promising potential in improving corrosion assessment methods. These techniques enable accurate prediction, early detection, and effective monitoring of corrosion in reinforced concrete (RC) structures. By leveraging data driven models, engineers can enhance the durability and resilience of infrastructure, minimizing maintenance costs and preventing catastrophic failures. Moving forward, adopting innovative strategies and materials resistant to corrosion will play a crucial role in ensuring the long-term sustainability of critical structures.

Chapter 02:

Stat of the Art and Corrosion Prediction Models

1. Introduction

Corrosion is one of the most persistent and costly challenges in the oil and gas industry, threatening the integrity and longevity of pipeline infrastructure. As pipelines age and operate in increasingly harsh and variable environments, accurate prediction of corrosion progression becomes vital for ensuring safe operation and minimizing economic losses. Traditional corrosion prediction models such as physical or statistical approaches have long been applied, but their limited adaptability to noisy, non-linear, and high dimensional data restricts their performance in modern industrial contexts.

In recent years, the rise of artificial intelligence (AI), and particularly machine learning (ML) and deep learning (DL), has offered transformative potential in this domain. These models are capable of learning complex relationships in historical inspection data and providing accurate forecasts of corrosion evolution over time. This chapter provides a comprehensive overview of corrosion prediction models starting from the foundations of physical and statistical methods, to the latest developments in ML and DL techniques. It also presents a structured review of scientific works that illustrate how these models have been implemented across various scenarios, highlighting their methodologies, datasets, evaluation metrics, and predictive performances.

2. Historique of Prediction Methods in Oil and Gas Industry

The history of corrosion prediction methods has evolved significantly over the decades. In the 1940s to 1990s, traditional physical models were dominant, focusing on life prediction (Schneider, 1947), corrosion rate estimation (Mahato, 1968), and fracture prediction (Schulze, 1980). Later, statistical methods emerged in the 1990s, using data-driven techniques for evaluating corrosion rates (Adams, 1993), soil corrosion (Jing, 1998), and stress analysis (Li et al., 2004).

From the 2000s onward, machine learning (ML) methods began to play a significant role, enabling more accurate and scalable solutions. These included models for corrosion degree evaluation (Li, 1998), defect assessment (Silva, 2007), failure pressure prediction (Khalajestani, 2014), pit corrosion prediction (Roy, 2018), and corrosion image classification (Bastian, 2019). Today, ML and deep learning techniques like BiLSTM and CNN are being increasingly adopted for precise and real-time corrosion forecasting.

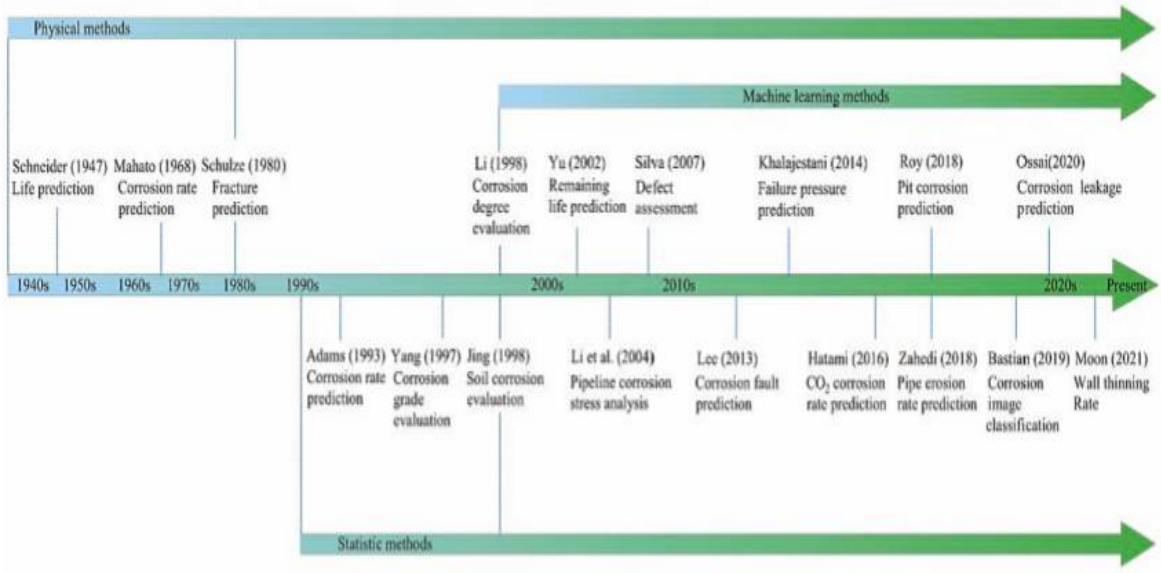


Figure 2.1: First application time and brief development history of the three methods

(Xu et al., 2023)

3. Physical Models

As early as 1947, Schneider applied the physical method to predict the lifespan of pipelines in the context of corrosion. This approach is primarily grounded in thermodynamic principles and the use of three fundamental conservation equations to analyze corrosion processes. Over the past seventy years, physical models have been extensively applied in the oil and gas industry for corrosion assessment. Typically, these models require detailed input parameters and provide reliable, practical predictions. However, they are highly dependent on numerous physical variables, which results in increased modeling complexity, high computational costs, and low efficiency. Additionally, such models often struggle with convergence and have limited scalability or adaptability to other scenarios (L. Xu et al., 2023a).

4. Statistical Models

In 1993, statistical methods were first introduced to predict the corrosion rate of oil pipelines. These methods are well-suited for developing predictive models based on historical data. Compared to other forecasting approaches, statistical models offer better real-time performance. Common techniques include linear regression, autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), the Grey model, Kalman filtering, and Markov chains. However, statistical approaches often suffer from limitations such as poor extrapolation ability, a narrow prediction range, and a high dependency on data quality. As a result, they are more appropriate for linear datasets, and their predictive accuracy may fall short in complex or nonlinear scenarios(L. Xu et al., 2023a).

5. Machine Learning Models

Corrosion poses a significant risk to oil and gas pipelines operating in harsh environments, making accurate prediction essential for safe and efficient service. While traditional empirical and mechanism-based models have been used with some success, their complexity and limited adaptability hinder broader application. In contrast, data-driven approaches, particularly those using machine learning (ML), offer improved accuracy and flexibility. This review explores the evolution, strengths, and limitations of ML-based corrosion prediction models, and presents a comprehensive database summarizing key elements such as preprocessing methods, input/output

parameters, and performance indicators providing valuable insights for model selection and future research directions (Q. Wang et al., 2023).

The machine learning models employed in **Figure 2.2** are shown, as are the expected outputs based on the relevant findings (Hussein Khalaf et al., 2024).

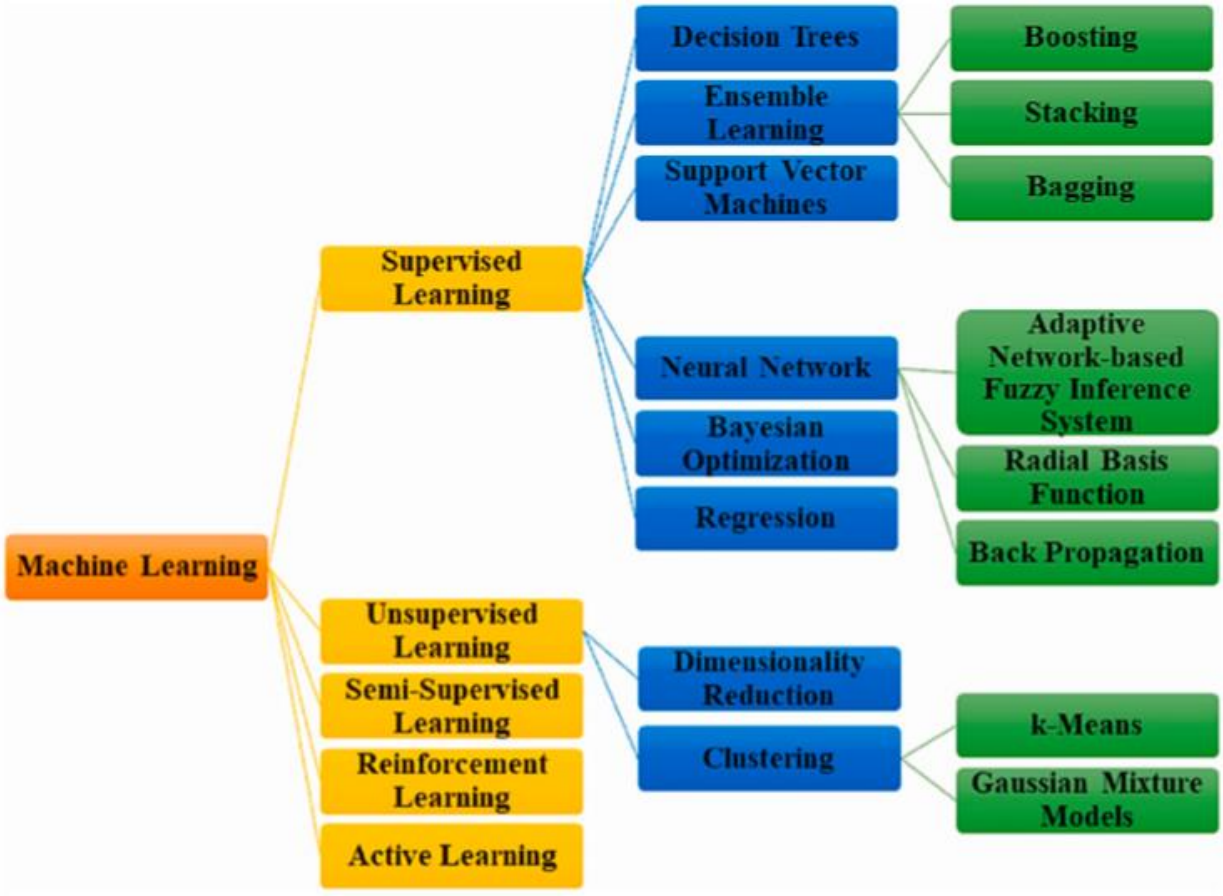


Figure 2.2: Machine learning models applied for corroded oil and gas pipelines integrity (Hussein Khalaf et al., 2024)

are shown in **Figure 2.3**. It is evident that the application of SNN and SVM accounts for a large proportion, followed by RF and DNN. Therefore, based on the current development of machine learning methods and by combining the business characteristics of each production link in the field of oil and gas pipeline corrosion, the typical applications of four kinds of machine learning methods in this field, including SNN, SVM, RF, and DNN are focused on in this pipeline.

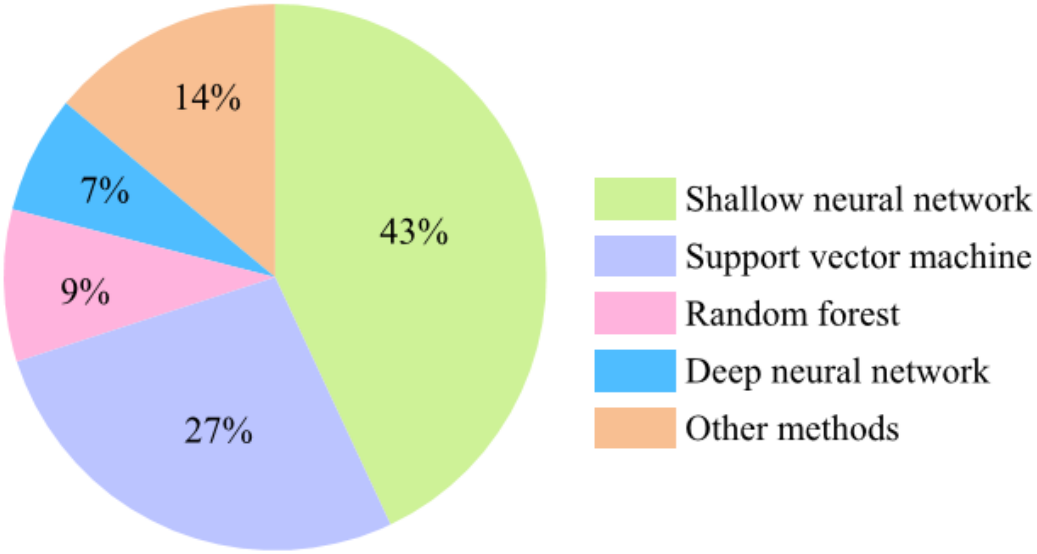


Figure 2.3: Statistics of benchmark models in the field of oil and gas pipeline corrosion prediction.

(Xu et al., 2023).

5.1 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANNs) are inspired by the human brain's structure and function, capable of learning complex nonlinear relationships through self-training and adaptive behavior. They consist of interconnected layers input, hidden, and output as shown in **Figure.2.4**,where neurons apply nonlinear functions to extract patterns from data. ANNs have been widely applied in corrosion prediction due to their high accuracy, especially models like Back-Propagation (BP) and Feed-Forward networks. Studies show that incorporating more input variables significantly improves their predictive performance, making ANNs highly effective for modeling corrosion behavior in oil and gas pipelines (Ma et al., 2023).

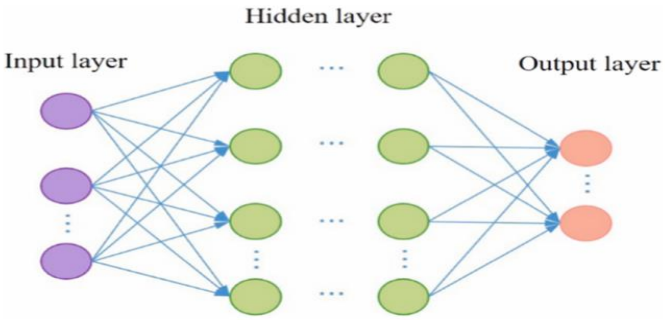


Figure 2.4: The artificial neural network architecture.

(Ogunbo et al., 2020)

5.2 Support Vector Machine (SVM)

In 2013,(Luo et al., 2013). applied Support Vector Machine (SVM) to predict the corrosion rate in offshore natural gas pipelines. The model incorporated a wide range of input parameters, including angle, pressure, deposition rate, fluid and gas densities, liquid velocity, pH, surface tension, flow regime, temperature, wall shear stress, and various thermal properties. The SVM model achieved a maximum relative error of 25% and an average error of 11.16%, outperforming both neural networks (maximum error of 33.33%, average of 19.54%) and multivariate analysis methods (maximum error of 44.44%, average of 25.32%). These results demonstrate the superior predictive performance of SVM in this context. In the same year, Lee and Rajkumar(Lee et al., 2013) employed a classification approach using Euclidean-based SVM in MATLAB to predict pipeline failure through long-range ultrasonic transducers (LRUT). Their findings revealed that the Euclidean-SVM classifier performed better than conventional SVM in identifying corrosion defects using LRUT data(Soomro et al., 2022a).

5.3 Random Forest (RF)

Random Forest (RF) is a powerful supervised ensemble learning method based on the concept of Classification and Regression Trees (CART). Initially proposed by Ho in 1995 (Ho, 1995) and later developed by Breiman in 2001(Biau& Scornet, 2016), RF builds a collection of decision trees to enhance prediction accuracy. The method operates in two phases: first, it generates multiple decision trees by randomly selecting subsets of training data and features using a bootstrap sampling technique; second, it aggregates the predictions from all trees to produce the final output typically by averaging (for regression) or majority voting (for classification). The accuracy of RF generally improves as the number of trees increases. Typically, two-thirds of the dataset are used for training, while the remaining one-third excluded via bootstrapping is used to validate the model's performance, ensuring robustness and generalization(Ben Seghier et al., 2022).

5.4 Decision Tree

The decision tree classifier is a hierarchical data structure that uses a divide and conquer strategy to classify data based on multiple attributes. In this study, the dataset includes features such as flow rate, temperature, pressure, and corrosion level (high or low). Since high corrosion is often linked to gas leakages, the decision tree model is employed to predict such leakages. The

tree begins with a root node and progressively splits into child nodes by asking questions based on input features. This process continues until each data point is classified into a specific category (Madamanchi et al., 2024a).

As illustrated in the **Figure2.5**, features like color and shape (e.g., circle, square, triangle) guide the decision-making at each node for example, asking "Is the color blue?" Each branch leads to a final decision or class label, such as A, B, or C. One of the main advantages of using a decision tree model is its ability to capture nonlinear relationships between predictors, such as temperature, flow rate, and pressure something linear models often cannot achieve. Therefore, decision trees are highly effective for analyzing complex patterns in pipeline corrosion and predicting potential leakages(Ben Zhou & C. Cervantes, 2016).

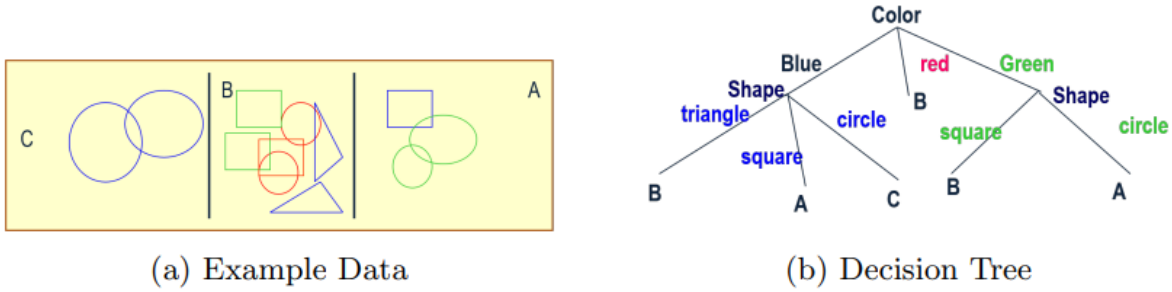


Figure 2.5: Decision Tree Example.
(Ben Zhou & C. Cervantes, 2016)

6. Deep Learning Models

Deep learning (DL), a subfield of machine learning, has gained considerable attention in recent years due to its exceptional ability to model complex, non-linear relationships in high-dimensional data. Unlike traditional models, DL architectures such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks can automatically extract hierarchical features from raw input data, making them highly suitable for corrosion prediction in challenging environments like oil and gas pipelines.

Recent studies have demonstrated the superior performance of DL in tasks such as corrosion detection, classification, and rate estimation. For example, (Li et al., 2022) developed an SSA-LSTM-based model for predicting the maximum pitting corrosion depth in subsea oil pipelines, achieving high accuracy and robustness in time-series prediction scenarios. Their findings

underscore the potential of DL to enhance predictive maintenance and risk assessment in industrial systems.

6.1 Deep Neural Network (DNN)

With the evolution of pipeline systems into intelligent networks, the resulting massive, high-dimensional data has challenged traditional machine learning methods, which struggle with complexity and limited ability to model non-linear functions. To overcome these issues, Deep Neural Networks (DNN) have gained prominence, supported by increased data availability and computing power. DNN approaches include supervised methods like RNN and CNN, and unsupervised methods such as DBN, GAN, and Auto-encoder (L. Xu et al., 2023a).

6.2 Recurrent Neural Network (RNN)

Is a type of artificial neural network specifically designed to handle sequential data by maintaining a memory of previous inputs through recurrent connections. Unlike traditional feed forward neural networks, RNNs have loops in their architecture, allowing information to persist and be passed from one step of the sequence to the next.

This study presents a recurrent neural network (RNN) based method for predicting the lifespan of equipment subjected to corrosion. The model utilizes time-series data to estimate the degradation process, providing a more accurate prediction of equipment life under corrosive conditions. The RNN approach captures the temporal dependencies inherent in corrosion progression, offering improved performance over traditional predictive models(Du et al., 2024).

6.3 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is an advanced type of Recurrent Neural Network (RNN) specifically designed to capture long-term dependencies in sequential data. Unlike traditional RNNs, which suffer from vanishing or exploding gradient problems when processing long sequences, LSTM networks use a unique memory cell architecture that allows them to retain relevant information over extended time intervals. This makes LSTM especially effective for modeling time-series data where the current state depends on historical patterns.

In this study, BiLSTMis used as the main component of the model to capture bidirectional temporal dependencies in pipeline corrosion time series. Unlike conventional LSTMs, BiLSTM

processes data in both directions (forward and backward), which can better exploit the past and future context at each time point in the sequence(J. Wang et al., 2024a).

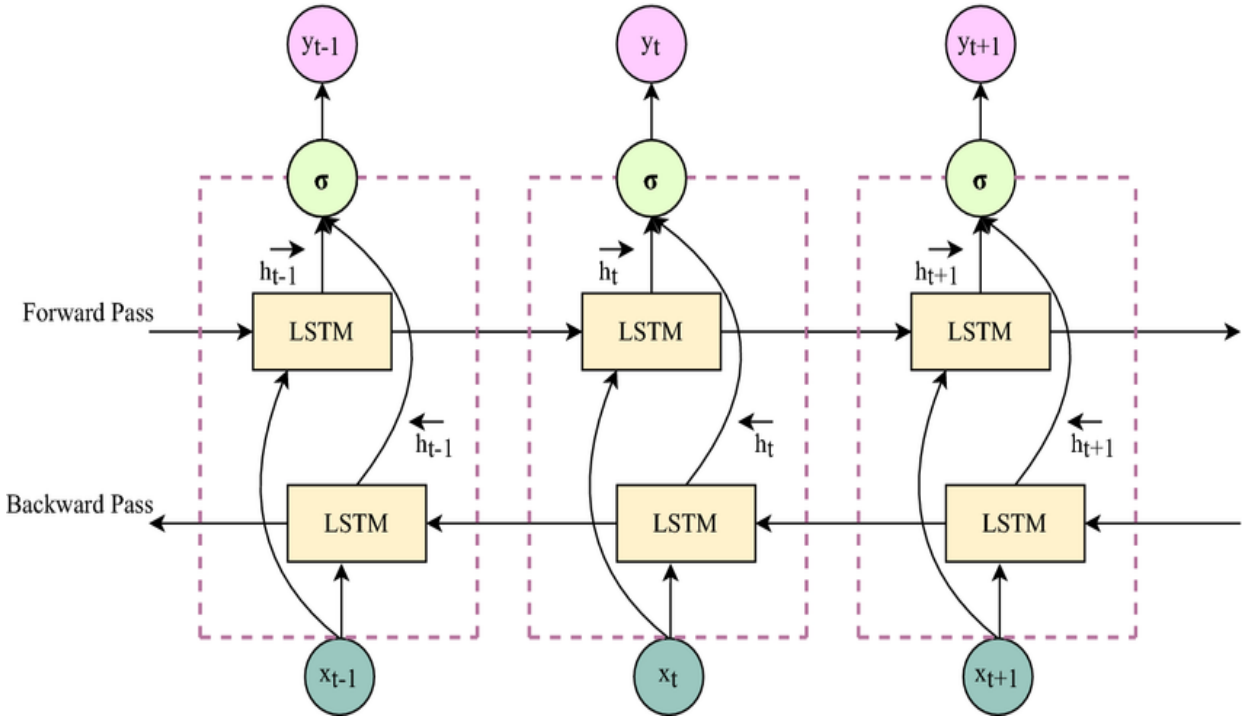


Figure 2.6: Bidirectional LSTM architecture (Dinesh Naik& C. D. Jaidhar, 2022)

6.4 Temporal Convolutional Networks

Temporal Convolutional Networks (TCNs) are a class of deep learning models designed for sequential data, which use 1D causal and dilated convolutions to capture temporal dependencies. Unlike traditional recurrent models like RNNs or LSTMs, TCNs rely entirely on convolutional architectures, allowing for parallel processing, stable gradients, and larger receptive fields. TCNs are particularly effective in time-series forecasting, including applications like corrosion prediction, due to their ability to model long-term dependencies with less computational cost (Lara-Benítez et al., 2020).

7. Further Work and Research Gaps

In recent years, numerous machine learning (ML) and deep learning (DL) techniques have been applied to improve corrosion prediction in the oil and gas industry. This section provides a review of relevant contributions, followed by a comparative summary and discussion of their significance and future directions.

7.1 Recent ML and DL Approaches for Corrosion Prediction in Oil and Gas Pipelines

The following studies illustrate the diversity of data-driven approaches:

- In (Imran et al., 2024),the authors compared Artificial Neural Networks (ANN) and Random Forests (RF) for corrosion prediction in maritime and offshore structures.ANN captured complex patterns, while RF offered interpretability and robustness. The study emphasized the need for quality data and hybrid approaches.
- The study by (Fang et al., 2023) used simulated data from OLI Studio Corrosion Analyzer to train Random Forest and CatBoost models for predicting internal corrosion. The models achieve high accuracy ($R^2 > 0.97$), outperforming traditional approaches. CatBoost emonstrated superior performance in handling nonlinear interactions within complex pipeline environments.
- In the study by (J. Wang et al., 2024b), the authors proposed a hybrid model combining BiLSTM-GRU, RF, and the Improved Binary Whale Optimization (IBWO) algorithm. The model accurately predicted corrosion using time-series variables such as pH and resistivity, achieving a high performance with $R^2 = 0.9876$.
- The study by (Canonaco et al. 2022) introduces a transfer learning-based approach for corrosion prediction in pipeline infrastructures with limited labeled data. By leveraging learned features from related domains, the neural network model significantly improves prediction accuracy, particularly under data-scarce conditions.
- In their study (Malashin et al. , 2024) introduced a convolutional neural network (CNN)-based computer vision system for detecting pitting corrosion in gas pipelines. Utilizing a dataset of 576,000 images, the custom-designed CNN achieved a high binary classification accuracy of 98.44%, effectively eliminating the need for manual inspection.
- In (C. Wang et al., 2019), the authors applied an Immune Algorithm to optimize a Support Vector Machine (SVM) model, resulting in the IA-SVM, for predicting corrosion in subsea

pipelines. This optimization improved hyper parameter tuning and enhanced prediction stability under varying environmental conditions.

- In their study,(Son et al., 2024) developed a machine learning model combining Mask R-CNN for corrosion area detection and CNN regression for depth prediction in marine structures, particularly ship hulls. Utilizing over 35,000 images from public sources and salt spray corrosion experiments, the model achieved a pixel accuracy of 94.2% and a mean absolute error (MAE) of 0.2565 mm in depth prediction. This research introduces an integrated corrosion monitoring system, offering significant advancements in ship maintenance by enabling early intervention and quantitative assessment.
- The study by (Madamanchi et al., 2024b)explored the application of machine learning algorithms to predict corrosion severity in oil and gas pipelines. The authors evaluated and compared three classification models: K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF), using an open-source dataset comprising over 8,000 records with variables such as pressure, flow rate, temperature, and actual corrosion levels.

  The Random Forest model outperformed the others, achieving an accuracy of 91.55%, a sensitivity of 92%, and a precision of 92%. The Decision Tree model followed with an accuracy of 87.13%, while the KNN model achieved 84% accuracy. These results indicate that RF provides the highest precision and robustness, but KNN remains a useful alternative due to its simplicity and efficiency in moderately complex environments.

- conducted by (Li et al., 2022) , introduced a hybrid Singular Spectrum Analysis (SSA) and Long Short-Term Memory (LSTM) model for predicting the maximum pitting corrosion depth in subsea oil pipelines. The model utilized time-series data such as temperature, pressure, flow rate, and various corrosion-related factors. The SSA component helped denoise and decompose the complex signals before feeding them into the LSTM, enhancing learning of temporal patterns. The SSA-LSTM model achieved a Mean Absolute Error (MAE) of approximately 0.052 mm and a Root Mean Squared Error (RMSE) of around 0.09 mm, significantly outperforming standalone LSTM and ARIMA models. This demonstrates the model's strong robustness and accuracy for corrosion prediction in dynamic subsea environments.

- published by (Yarveisy et al., 2024), focused on predicting the burst pressure of corroded pipelines using a data-driven regression-based machine learning approach. The model used pipeline inspection data, including wall thickness, defect dimensions, and defect location. After testing multiple algorithms, the Gradient Boosting Regressor provided the best performance, achieving an R^2 score of 0.96, a Mean Absolute Error (MAE) of 5.3 MPa, and a Root Mean Squared Error (RMSE) of 6.8 MPa. These results enabled accurate assessment of residual pipeline strength, contributing to more reliable pipeline safety evaluations and maintenance planning

7.2 Summary Table

Table 2.1 summarizes the key characteristics of the above studies, including:

- Type of model (ML, DL, or hybrid)
- Input parameters (e.g: pH, temperature, ILI data, images)
- Predicted outputs (e.g: corrosion depth, burst pressure)
- Main results (accuracy, performance metrics, innovations)

Table 2.1: Summary of Research Works Using ML/DL for Corrosion Prediction in the Oil and Gas Industry

Title of Article	Year	Authors	ML/DL Method Used	Type	Input Parameters	Prediction Output	Key Results
A critical review of ML algorithms in maritime, offshore, and oil & gas corrosion research: A comprehensive analysis of ANN and RF models	2024	(Imran et al., 2024)	ANN, RF	Machine Learning	Temperature, humidity, chemical exposure	Corrosion behavior and severity	Comparative study; ANN effective but complex; RF more interpretable
Development of ML algorithms for predicting internal corrosion of crude oil and natural gas pipelines	2023	(Fang et al., 2023)	RF, CatBoost	Machine leaning	Pressure, flow, chemical concentration	Internal corrosion rate (mm/year)	High accuracy; CatBoost robust with complex input; RF achieves $R^2 > 0.97$
Corrosion Rate Prediction for Buried Oil and Gas Pipelines Using RF and IBWO-Optimized BiLSTM-GRU Model	2024	(J. Wang et al., 2024b)	BiLSTM-GRU + RF + IBWO	**Hybride**(ML + DL)	Time-series: pH, soil temp, resistivity	Corrosion rate progression	$R^2 = 0.9876$, MSE = 0.0498; high performance in dynamic environments
A Transfer-Learning Approach for Corrosion Prediction in Pipeline Infrastructures	2022	(Canonaco et al., 2022)	Transfer Learning, Neural Networks	Deep Learning	Pipeline specs, simulated corrosion data	Corrosion rate	Improves prediction with limited labeled data using cross-domain knowledge
Deep Learning Approach for Pitting Corrosion Detection in Gas Pipelines	2024	(Malashin et al., 2024)	Convolutional Neural Network (CNN)	Deep Learning	Images of pipeline surface	Corrosion presence (binary)	CNN achieved 98.44% accuracy in detecting pitting corrosion
Prediction of Corrosion Rate of Submarine Oil and Gas Pipelines Using IA-SVM Model	2019	(C. Wang et al., 2019)	Immune Algorithm + SVM	Machine Learning	Temperature, pH, chloride concentration	Corrosion rate	IA optimization improves SVM performance with lower prediction errors

Corrosion Area Detection and Depth Prediction Using Machine Learning	2024	(Son et al., 2024)	Mask R-CNN for segmentation + CNN regression	Deep Learning	Corrosion images, color maps, illumination data	Corrosion area + Depth estimation	Detection accuracy 94.2%, MAE for depth as low as 0.2565 mm with JET color model
A Machine Learning-Based Corrosion Level Prediction in the Oil and Gas Industry	2024	(Madamanchi et al., 2024b)	K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF)	Machine Learning	Sensor data: pressure, temperature, flow rate, corrosion level	Corrosion level classification	KNN achieved 84% accuracy; RF outperformed with 91.55%; suitable for small datasets
A data-driven prediction model for maximum pitting corrosion depth of subsea oil pipelines using the integrated SSA and LSTM approach	2022	(Li et al., 2022)	SSA + LSTM	Deep Learning	Time series data: temp., pressure, flow rate, corrosion factors	Max pitting corrosion depth	SSA-LSTM outperforms LSTM and ARIMA. MAE ≈ 0.052 mm, RMSE ≈ 0.09 mm. Demonstrates high robustness and accuracy on time-series corrosion data.
Data-Driven Model to Predict Burst Pressure in the Presence of Corrosion Defects	2024	(Yarveisy et al., 2024)	Data-Driven Modeling (Regression Approach)	Machine Learning	Pipeline inspection data (wall thickness, defect size, location)	Burst pressure of corroded pipelines	Gradient Boosting Regressor gave best results. $R^2 = 0.96$, MAE = 5.3 MPa, RMSE = 6.8 MPa. Effective for residual strength estimation and safety assessment.

This comparative view reveals the effectiveness and applicability of modern ML/DL approaches across diverse corrosion scenarios.

7.3 Discussion

The review of recent research clearly indicates that machine learning (ML) and deep learning (DL) techniques are playing an increasingly vital role in corrosion prediction for oil and gas pipeline systems. Various ML models, such as Random Forest (RF), Support Vector Machines (SVM), and CatBoost, have demonstrated high predictive accuracy, particularly when dealing with structured sensor data and operational parameters. RF emerged as one of the most widely adopted methods, offering strong interpretability and robustness across several studies, including

those by (Fang et al., 2023) and (Madamanchi et al., 2024b). In parallel, deep learning models, particularly CNN and BiLSTM architectures, have proven highly effective in image-based corrosion detection and time-series forecasting tasks. For instance, (Malashin et al., 2024)achieved over 98% accuracy in pitting corrosion classification using CNN, while (J. Wang et al., 2024b)combined BiLSTM-GRU and optimization techniques to enhance time-series predictions with excellent performance ($R^2 > 0.98$).

Hybrid models integrating ML and DL techniques are also gaining traction, as they leverage the strengths of both approaches. The hybrid ANN-SVM models reviewed by (Soomro et al., 2022b)and the SSA-LSTM model by (Li et al., 2022)demonstrated superior accuracy and resilience to data variability. Moreover, transfer learning and data augmentation strategies have been proposed to address the issue of limited labeled data, enhancing generalization in real-world scenarios.

Despite these advancements, several challenges remain. These include model explainability, the availability of high-quality and labeled inspection datasets, and the adaptation of models to different pipeline environments. Furthermore, the review suggests a strong need for integrated corrosion monitoring systems that combine spatial tracking, automated defect recognition, and predictive modeling. Overall, the evolving landscape highlights the transition from empirical approaches to data-driven intelligence, paving the way for more proactive and precise pipeline integrity management.

8. Conclusion

This chapter has presented a comprehensive overview of the existing corrosion prediction models used in the oil and gas industry, ranging from traditional physical and statistical approaches to modern machine learning (ML) and deep learning (DL) techniques. While physical and statistical models have historically laid the foundation for corrosion assessment, their limitations in handling complex, nonlinear, and high-dimensional data have led to the emergence of data-driven approaches. ML and DL models such as Random Forest, SVM, CNN, and BiLSTM have demonstrated significant potential in improving prediction accuracy, automating feature detection, and enabling real-time monitoring. Furthermore, hybrid and transfer learning methods have emerged as promising directions to overcome challenges related

to data scarcity and model generalization. The insights from this review not only underline the growing impact of intelligent algorithms in predictive maintenance but also highlight current gaps and future opportunities for advancing corrosion monitoring technologies in industrial applications.

Chapter 03:

Implementation and Evaluation of the Prediction Model

1. Introduction

To develop a reliable corrosion prediction model for oil and gas pipelines, it is essential to start with high-quality, well-processed data. This chapter first presents a detailed description of the dataset, including its origin, structure, and key environmental and operational variables influencing corrosion.

Following the dataset overview, a thorough data analysis is conducted to ensure data consistency and suitability for deep learning modeling. This process involves handling missing values, detecting and correcting outliers, and studying the relationships between variables using correlation matrices.

These data preparation steps are critical to optimize the performance of the deep learning model, particularly the Long Short-Term Memory (LSTM) network employed in this project. The LSTM architecture, known for its ability to learn complex temporal patterns, requires clean and well-structured input data to accurately predict the evolution of corrosion over time. Thus, this chapter lays the necessary groundwork for building an efficient and robust LSTM-based corrosion prediction system.

2. Dataset Description

The dataset employed in this project originates from the "Cross-country Pipeline Inspection Data Analysis and Testing of Probabilistic Degradation Models"published by RiosharYarveisy, Faisal Khan, and RouzbehAbbassi (2021) (Yarveisy et al., 2021) study and focuses on the structural degradation of oil and gas pipelines. Specifically, the dataset includes inspection records from four key years: Year 1, Year 3, Year 5, and Year 7, which represent sequential ILI runs conducted at regular two-year intervals over a seven-year span. This temporal structuring allows for the tracking and modeling of corrosion growth over time, enabling longitudinal analysis of individual corrosion anomalies across multiple inspection periods.

The dataset contains approximately 15 variables, with features classified into numerical and categorical data types. The main variables are:

- **Year**: Inspection year (note: does not reflect actual pipeline age) – Numeric/Categorical

- **GWNUM**: Girth weld number associated with the pipeline section(Number of segment , It begins 14) – Numeric/Categorical
- **JL.m**: Joint length of the pipeline segment in meters – Numeric
- **RD.m:** Feature's relative location to the pipe joint (in meters) – Numeric (Corrosion location from segment)
- **SO.deg**: Pipe joint's longitudinal seam weld orientation (degrees) – Numeric (missing for year three data)
- **SPD.m**: Absolute distance of the feature's starting point from the pipeline origin (meters) – Numeric (Corrosion location from pipeline)
- **SPO.deg**: Starting point circumferential location (degrees) – Numeric
- **EPD.m**: Absolute distance of the feature's endpoint from the pipeline origin (meters) – Numeric (missing for year five data)
- **EPO.m**: Endpoint circumferential location (degrees) – Numeric
- **SIPRD.m**: Most significant point relative location to the pipe joint (degrees) – Numeric **(**Corrosion Center**)**
- **SIPO.deg**: Most significant point circumferential location (degrees) – Numeric
- **WT.mm**: Nominal wall thickness (millimeters) – Numeric/Categorical
- **L.mm**: Feature length (millimeters) – Numeric
- **W.mm**: Feature width (millimeters) – Numeric
- **MD.mm**: Feature maximum depth (millimeters) – Numeric

These variables provide detailed spatial, geometric, and structural information essential for understanding corrosion characteristics. Notably, attributes such as WT.mm and MD.mm are critical for evaluating corrosion severity and progression.

Given its temporal depth and spatial precision, this dataset is particularly well-suited for time-series modeling using deep learning architectures like Long Short-Term Memory (LSTM) or Bidirectional LSTM, which can effectively capture degradation trends and support predictive maintenance strategies.

3. Framework and Implementation

The implementation process is structured around a three-step pipeline to effectively deploy a BiLSTM based deep learning model for corrosion depth prediction. As illustrated in **Figure 3.1**, the workflow begins with **Step 01**: Data Preparation, which includes collecting pipeline inspection data from years 1, 3, 5, and 7, followed by cleaning, preprocessing, and feature matching across inspections. The combined dataset is then split into training (80%) and testing (20%) subsets. In **Step 02**: Model Implementation, a BiLSTM architecture is trained using temporal sequences extracted from the preprocessed data to capture degradation patterns. Finally, **Step 03**: Results Evaluation assesses the model's predictive performance using metrics such as MAE and MSE, along with visual comparison between predicted and actual corrosion depths.

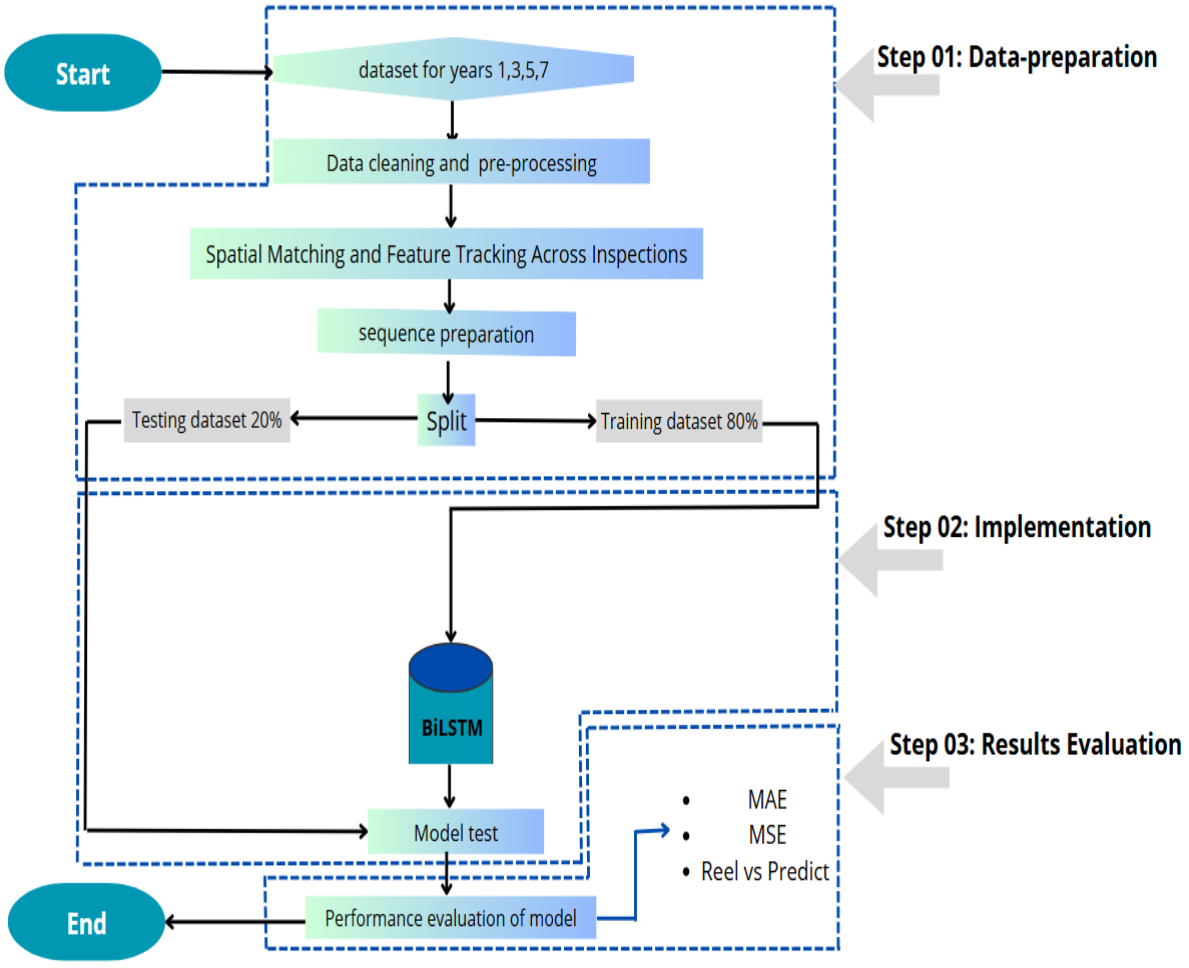


Figure 3.1: Framework for Corrosion Prediction Using BiLSTM

4. Data Cleaning

In order to ensure the reliability and quality of the dataset before applying deep learning model LSTM, several preprocessing and analysis steps are performed. This phase includes the treatment of missing values and correlation analysis among features. These operations aim to enhance the dataset consistency and improve the predictive performance of the model.

4.1 Handling Missing Values

The aim of this step is to identify and quantify the variables with missing values in the dataset. A clear understanding of missing data is crucial to ensure the quality and reliability of the predictive model (LSTM).

A Python script was used to calculate the percentage of missing values for each variable, in years 1,3,5, and 7.

Year 1:

No variables have missing values (0% missing for all columns).

```
Missing Values for Year 1 data:
Year        0.0
GWNUM       0.0
JL.m        0.0
RD.m        0.0
SO.deg      0.0
SPD.m       0.0
SPO.deg     0.0
EPD.m       0.0
EPO.deg     0.0
SIPRD.m     0.0
SIPO.deg    0.0
WT.mm       0.0
L.mm        0.0
W.mm        0.0
MD.mm       0.0
dtype: float64
```

Figure 3.2: Year1 -Missing values-.

Year 3:

The variable SO.deg shows 100% missing values.

```
Missing Values for Year 3 data:
Year           0.0
GWNUM          0.0
JL.m           0.0
RD.m           0.0
SO.deg       100.0
SPD.m          0.0
SPO.deg        0.0
EPD.m          0.0
EPO.deg        0.0
SIPRD.m        0.0
SIPO.deg       0.0
WT.mm          0.0
L.mm           0.0
W.mm           0.0
MD.mm          0.0
dtype: float64
```

Figure 3.3: Year3 -Missing values.

Year 5:

The variable EPD.m shows 100% missing values.

```
Missing Values for Year 5 data:
Year           0.0
GWNUM          0.0
JL.m           0.0
RD.m           0.0
SO.deg         0.0
SPD.m          0.0
SPO.deg        0.0
EPD.m        100.0
EPO.deg        0.0
SIPRD.m        0.0
SIPO.deg       0.0
WT.mm          0.0
L.mm           0.0
W.mm           0.0
MD.mm          0.0
dtype: float64
```

Figure 3.4: Year5 -Missing values-.

Year 7:

No variables have missing values (0% missing for all columns).

```
Missing Values for Year 7 data:
Year         0.0
GWNUM        0.0
JL.m         0.0
RD.m         0.0
SO.deg       0.0
SPD.m        0.0
SPO.deg      0.0
EPD.m        0.0
EPO.deg      0.0
SIPRD.m      0.0
SIPO.deg     0.0
WT.mm        0.0
L.mm         0.0
W.mm         0.0
MD.mm        0.0
dtype: float64
```

Figure 3.5: Year7 -Missing values-.

These results show that SO.deg is entirely missing in Year 3, and EPD.m is entirely missing in Year 5, while all other years contain complete data. As a result, these two variables will be excluded from the modeling process to avoid introducing inconsistencies.

4.2 Correlation Matrix

The correlation matrix is a fundamental statistical tool used to evaluate the linear relationships between variables in a dataset. Each element of the matrix represents the Pearson correlation coefficient between two variables, with values ranging from -1 to 1. A value close to 1 indicates a strong positive correlation, meaning that the two variables tend to increase together. Conversely, a value near -1 reflects a strong negative correlation, where one variable increases as the other decreases. A value around 0 suggests little to no linear relationship between the variables.

Year1:

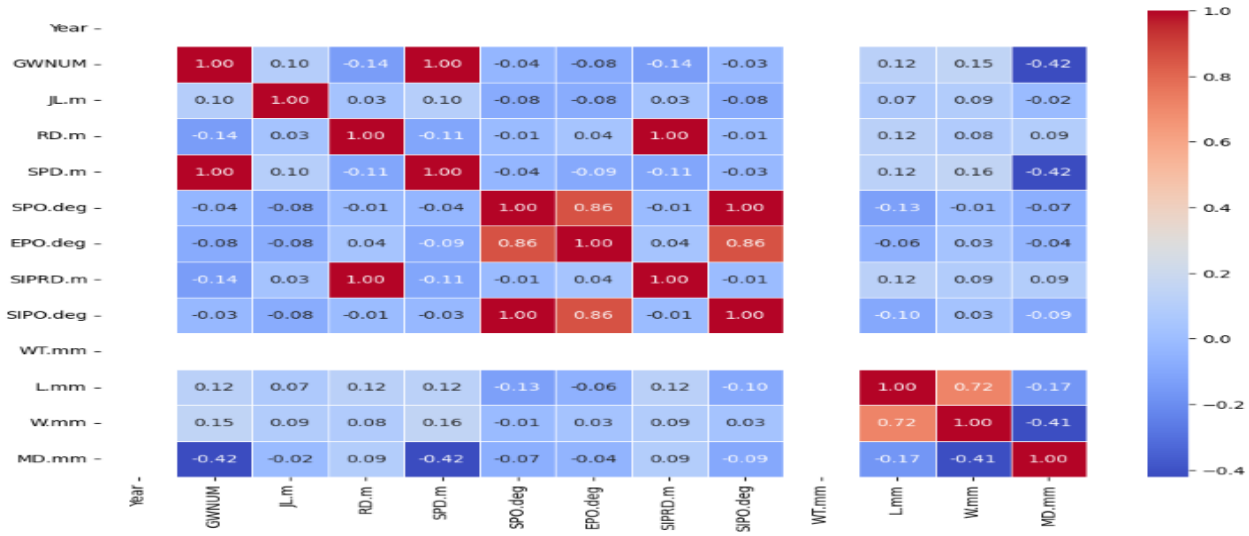


Figure 3.6: Correlation Matrix for year 1.

- Strong correlations are observed among the angular position variables:
  - SPO.deg, EPO.deg, and SIPO.deg show coefficients above 0.86, indicating possible redundancy in positional information.
- A perfect correlation (1.00) is found between GWNUM and SPD.m, suggesting that both may represent the same underlying variable.
- The target variable MD.mm (maximum corrosion depth) shows a moderate negative correlation with GWNUM (-0.42), possibly indicating early signs of temporal evolution in corrosion.
- Other variables like WT.mm, L.mm, and W.mm exhibit weak correlations with most parameters, reflecting limited impact in this early stage.

Year3:

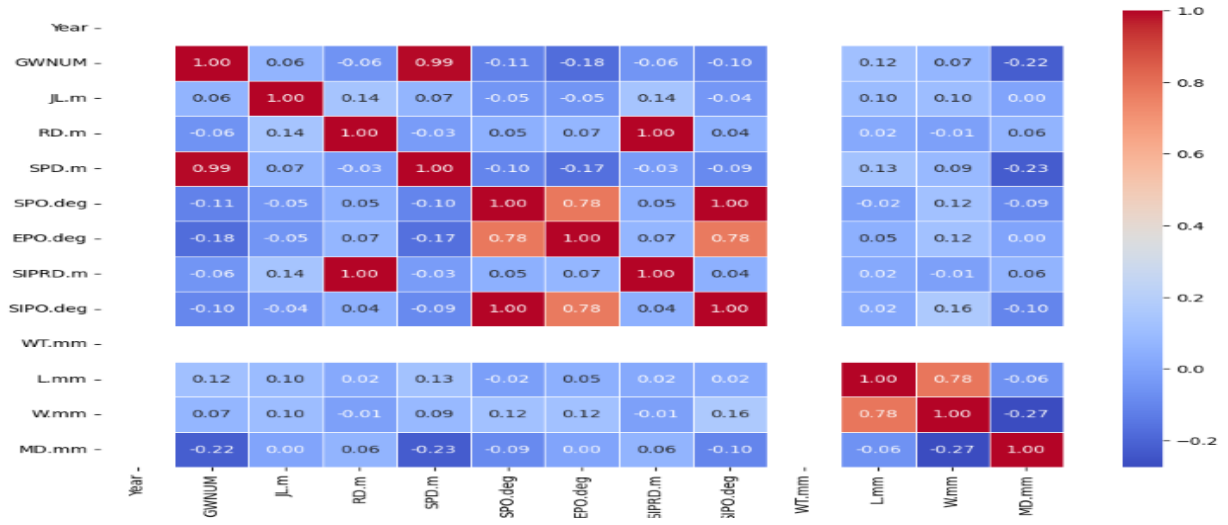


Figure 3.7: Correlation Matrix for year 3.

- Correlations between SPD.m and GWNUM (r ≈ 0.99), and among angular variables (up to 1.00) are still consistent, reinforcing the stability of spatial measurements.
- The correlation of MD.mm with positional variables like SPD.m and SPO.deg becomes more noticeable and negative, suggesting an increasing influence of spatial location on corrosion depth.
- A stronger correlation appears between L.mm and W.mm (r = 0.78), possibly reflecting the spatial expansion of corrosion over time

Year5:

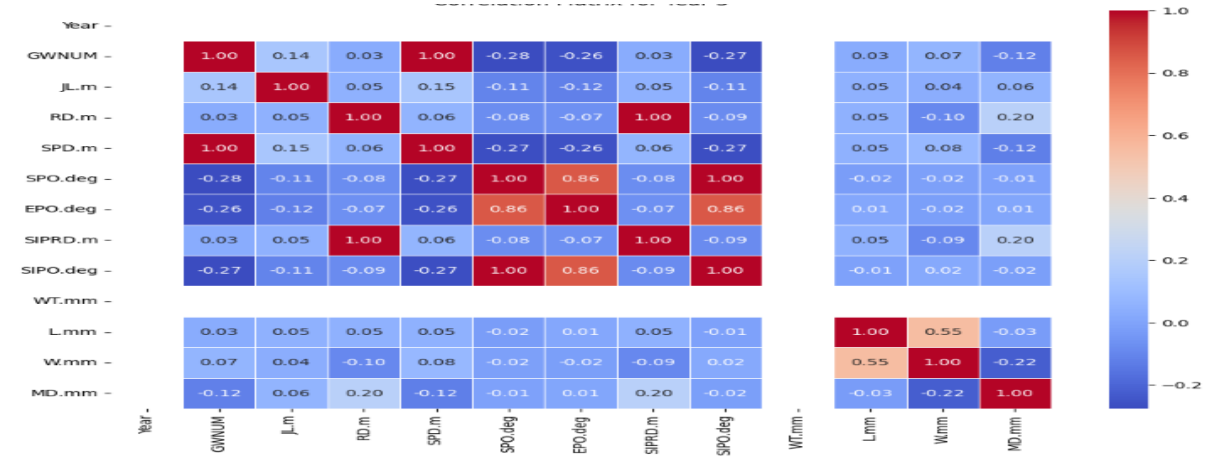


Figure 3.8: Correlation Matrix for year 5.

- Angular variables maintain high intercorrelations (~0.86), and SPD.m remains perfectly correlated with GWNUM (1.00).
- The correlations of MD.mm with other features remain low or negligible, though some slight negative correlations persist with SPD.m, GWNUM, and SPO.deg.
- This suggests that at this stage, corrosion depth may be influenced by external factors not captured in the current feature set, or its impact is still not strongly reflected in the available descriptors.

Year7:

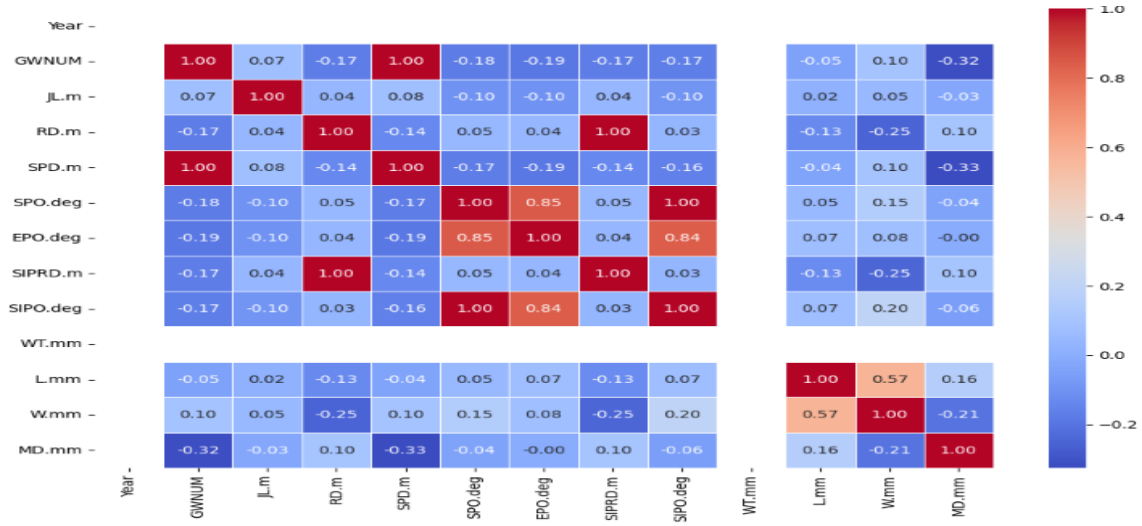


Figure 3.9: Correlation Matrix for year 7.

- High correlations persist between angular variables and between GWNUM and SPD.m.
- MD.mm now displays moderate negative correlations with:
  - SPD.m (-0.33)
  - GWNUM (-0.32)

  These trends support the hypothesis that corrosion depth increases with time and distance, reflecting progressive degradation of the pipeline material.
- The correlation between L.mm and W.mm remains noteworthy (r = 0.57), further supporting the idea of corrosion spreading both longitudinally and transversely within the pipeline.

The correlation study across Years 1, 3, 5, and 7 shows consistent high correlations between angular variables and confirms redundancy between GWNUM and SPD.m. The target variable

MD.mm reveals a **progressive negative correlation** with positional features over time, indicating that **corrosion depth increases with distance and time**. Additionally, the growing correlation between **L.mm and W.mm** suggests that corrosion spreads both **longitudinally and transversely**. These findings support the **temporal and spatial evolution** of corrosion and guide relevant feature selection for prediction models.

5. Spatial Matching and Feature Tracking Across Inspections

To track the progression of individual corrosion defects across several inspection campaigns (Years 1, 3, 5, and 7), a feature matching algorithm was implemented using the following steps:

- **Data Cleaning and Preparation:** The dataset initially included inspection records from four distinct time points: Year 1, Year 3, Year 5, and Year 7. During preprocessing, entries with missing values in key variables (such as SO.deg and EPD.m) were removed to ensure data integrity and reliability for the subsequent spatial matching and predictive modeling steps.
- **Standardization of Coordinate Format:** The spatial position columns SIPRD.m (relative longitudinal position) and SIPO.deg (circumferential position) were converted to numeric format, replacing commas with dots where necessary.
- **Initialization of Corrosion Identifiers:** A new column corrosion_id was created to uniquely identify each matched corrosion defect across the years.
- **Defining Matching Thresholds:** To determine whether two features are spatially close, we used:
  - Maximum difference of **15°** for circumferential position (SIPO.deg).
  - Maximum difference of **0.05 m** for longitudinal position (SIPRD.m).
- **Segment-wise Matching (by GWNUM):** Matching was performed independently for each girth weld segment (GWNUM), to ensure spatial consistency within pipe segments.
- **Sequential Year-by-Year Matching:** Starting from defects detected in Year 1, the algorithm attempts to find the closest matching defect in Year 3, then Year 5, and finally Year 7, based on the smallest Euclidean distance using SIPO.deg and SIPRD.m.
- **Assignment of Corrosion IDs:**If a consistent match is found across all four years, the same corrosion_id is assigned to the corresponding records in Years 1, 3, 5, and 7.

- **Validation of Complete Matches:** Only corrosion IDs that appear in all four inspection years are retained to ensure complete temporal tracking of defect evolution.
- **Final Dataset for Modeling:** The resulting dataset (df_matched) contains only matched defects with full 4-year histories. In total, over **781,784** matched records were identified.

Figure 3.10 below shows the distribution of features without explicit IDs, highlighting spatial clusters that persist across multiple years. This confirms that several corrosion spots appear consistently in the same area over time.

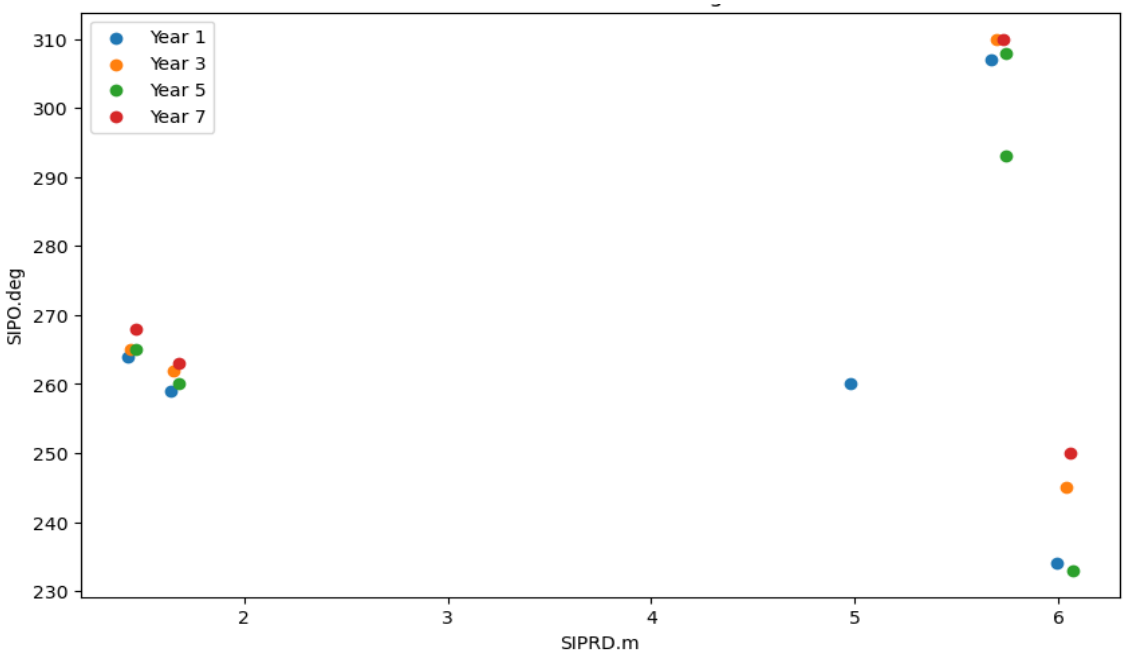


Figure 3.10: Distribution of SIPRD.m and SIPO.deg for GWNUM 22.

Figure 3.11 demonstrates spatial feature matching using assigned IDs (e.g., ID26, ID27, ID28), enabling better tracking of individual defects. These matches reflect the application of a geometric matching strategy based on relative distances and angles, which facilitates the longitudinal monitoring of specific corrosion features.

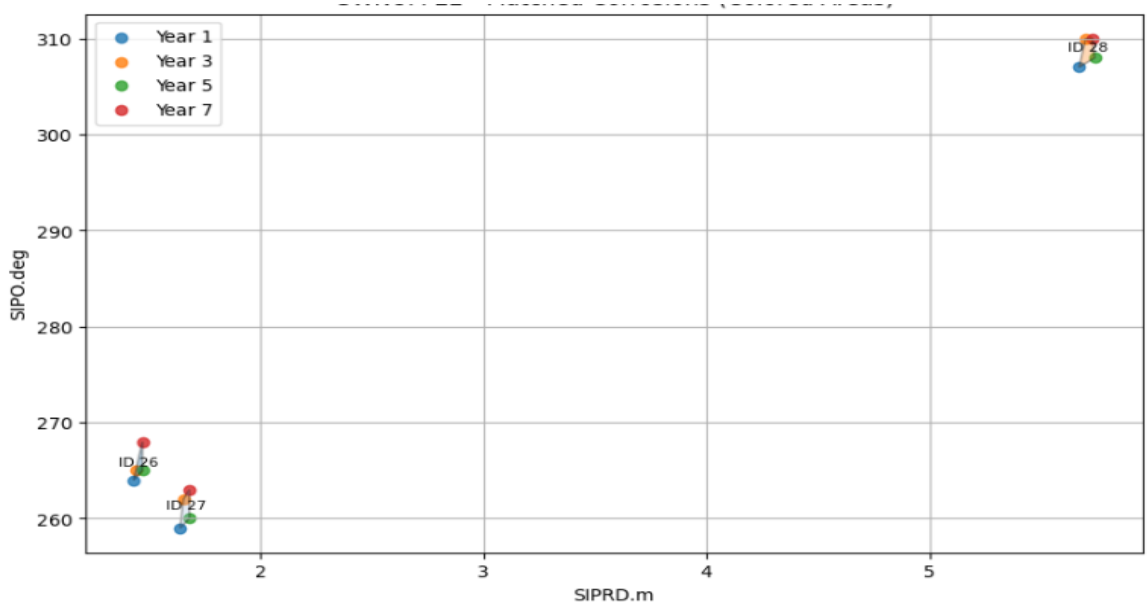


Figure 3.11: GWNUM 22 - Matched corrosions.

6. Implementation and Results

6.1 Bidirectional LSTM for Corrosion Depth Prediction

In this study, a Bidirectional Long Short-Term Memory (BiLSTM) model was developed to predict the maximum corrosion depth (MD.mm) at year 7 using time-series data from years 1, 3, and 5. The dataset includes pipeline inspection features such as joint length, radial distances, circumferential positions, and geometrical dimensions of corrosion defects.

6.2 Data Preparation

- The input features used were: ['L.mm', 'W.mm', 'SIPRD.m' , 'SIPO.deg','MD.mm'].
- The target variable was MD.mm, representing the corrosion depth at year 7.
- Each feature and target was scaled using MinMaxScaler to normalize the values between 0 and 1.
- Data was grouped by corrosion_id and sequences of inspections at years 1, 3, and 5 were used to predict the value at year 7.

6.3 Model Architecture

The BiLSTM model was structured as follows:

- Input Layer: Sequences of shape (3, 5) representing the three years and eight features.
- Bidirectional LSTM Layers: Two stacked BiLSTM layers with 64 units each to capture both past and future dependencies in the time-series.
- Dropout Layers: Dropout (rate = 0.3) was applied to reduce overfitting.
- Dense Layers: A fully connected layer with 32 units followed by an output layer with a single unit for regression.

Layer (type)	Output Shape	Param #
input_layer_12 (InputLayer)	(None, 3, 5)	0
bidirectional_24 (Bidirectional)	(None, 3, 128)	35,840
dropout_24 (Dropout)	(None, 3, 128)	0
bidirectional_25 (Bidirectional)	(None, 128)	98,816
dropout_25 (Dropout)	(None, 128)	0
dense_24 (Dense)	(None, 32)	4,128
dense_25 (Dense)	(None, 1)	33

Total params: 138,817 (542.25 KB)
Trainable params: 138,817 (542.25 KB)
Non-trainable params: 0 (0.00 B)

Figure 3.12: BiLSTM model structured.

6.4 Model summary

- Total parameters: 138,817
- Trainable parameters: 100%
- Non-trainable parameters: 0

6.5 Training Configuration

- Optimizer: Adam
- Loss function: Mean Squared Error (MSE)
- Metrics: Mean Absolute Error (MAE)
- Early stopping: Enabled to prevent overfitting based on validation loss.
- Batch size: 64.
- Epochs: 20

6.6 Model Performance

On the test set, the model achieved:

- Mean Absolute Error (MAE): ≈ **0.13**
- Mean Squared Error (MSE): ≈ **0.04**

6.7 Discussion of Prediction Results

The proposed BiLSTM-based deep learning model demonstrates promising performance in predicting the maximum corrosion depth (MD.mm) at year 7 using historical inspection data collected at years 1, 3, and 5. The matching algorithm developed to track individual corrosion defects across these years ensures temporal consistency and significantly enriches the dataset by linking features based on their spatial coordinates (SIPRD.m and SIPO.deg). A total of over 781,000 matched records were extracted, enabling the model to learn from long-term degradation patterns.

The scatter plot comparing predicted versus actual corrosion depths shows that the predictions generally align well with the ideal reference line (y = x), indicating high model accuracy. Minor deviations are observed, particularly in the upper value range (>2.5 mm). These discrepancies can be attributed to several real-world factors such as:

- Periodic maintenance actions (e.g: coating application or surface treatment).
- Noise or incompleteness in the historical data.
- Unrecorded changes in environmental conditions (e.g: pH, humidity, chemical exposure).

Moreover, the model's ability to generalize across diverse segments and inspection years showcases its robustness and adaptability to pipeline integrity analysis. The use of BiLSTM enables effective learning from time-series sequences, capturing both forward and backward temporal dependencies.

Overall, this methodology proves effective for predictive maintenance by enabling early detection of high-risk corrosion zones. It supports informed decision-making, reduces inspection costs, and enhances pipeline safety, offering a scalable solution for industrial corrosion monitoring.

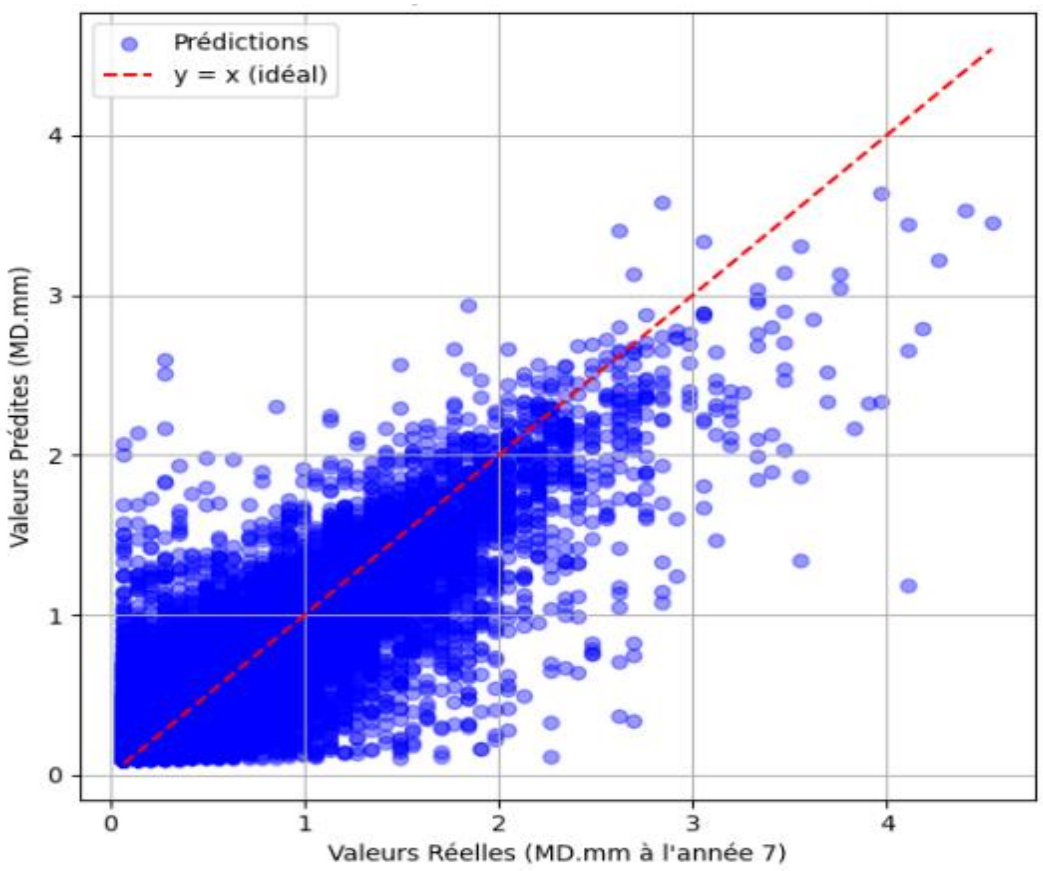


Figure 3.13: Predicted vs Actual Corrosion Depth (MD.mm) at Year 7.

7. Embedded System Using Raspberry Pi 4

To enable real-time, localized corrosion prediction in oil and gas pipelines, we propose an embedded deployment strategy built around the Raspberry Pi 4. This approach leverages edge computing to bring machine learning inference closer to the inspection site, reducing reliance on cloud infrastructure while maintaining efficiency, scalability, and cost-effectiveness.

7.1 Embedding the BiLSTM Model for Edge Inference

The core prediction model a BiLSTM network trained on historical inspection data is planned to be embedded into the Raspberry Pi 4. To meet the device's resource constraints, post training quantization techniques are applied, reducing model size and memory usage while maintaining acceptable accuracy. Once deployed, the Raspberry Pi can perform real-time inference on new inspection data, forecasting future corrosion depth without external computation.

7.2 Dataset Partitioning and External Storage

Given the size of the complete dataset (240 GB), it cannot be stored directly on the Raspberry Pi's internal memory. To address this limitation, the dataset is preprocessed and partitioned into manageable subsets, each containing data for a set of pipeline segments. These subsets are stored on external USB storage devices (e.g: SSDs or HDDs) connected to the Raspberry Pi. This structure allows the device to load and process data set by set, optimizing memory usage while enabling scalable and flexible deployment across different segments.

7.3 Optimized Spatial Matching and Feature Tracking per Segment

During corrosion tracking, the matching algorithm compares each feature only with historical features belonging to the dataset subset corresponding to the targeted pipeline segment. This focused comparison significantly reduces the number of features involved in the matching process, thereby lowering execution time and computational complexity.

7.4 Benefits of Raspberry Pi-Based Deployment

The proposed embedded deployment offers multiple operational benefits:

- **Real-Time Inference:** Enables immediate assessment of corrosion risk at the inspection site.
- **Efficient Processing:** Localized matching per segment reduces execution time and computation needs.
- **Resilience in Remote Environments:** Operates reliably in areas with limited connectivity.
- **Cost-Effective and Scalable:** Raspberry Pi 4 balances performance with affordability, allowing for large-scale distributed deployment.
- **Energy Efficiency:** Low power consumption makes it suitable for remote or energy-constrained field applications.

8 Conclusion

This chapter presented the practical implementation of the proposed corrosion prediction system for oil and gas pipelines, based on deep learning techniques. From data preparation and cleaning to spatial feature matching and model development, each step was carefully designed to ensure consistency and reliability. The Bidirectional LSTM (BiLSTM) model was trained using a refined and temporally aligned dataset, capturing the evolution of corrosion defects across multiple inspection years. The model achieved a high prediction accuracy, with low error rates and strong correlation between actual and predicted values. These results demonstrate the effectiveness of deep learning in modeling complex degradation phenomena and confirm the potential of our approach for real-time monitoring and predictive maintenance in industrial environments.

Beyond model development, this chapter also outlined a scalable and cost-effective embedded deployment strategy using Raspberry Pi 4 devices. By embedding the trained BiLSTM model and processing segmented inspection data locally, the system supports real-time inference, efficient feature matching, and segment-based corrosion tracking, all while operating under resource constraints typical of field environments. This edge-based architecture paves the way for distributed, intelligent monitoring and predictive maintenance systems in the oil and gas industry.

General Conclusion

This project presented the development of an embedded system for the analysis and prediction of corrosion in oil and gas pipelines, addressing a critical challenge in industrial infrastructure management. By leveraging deep learning techniques specifically the Bidirectional Long Short-Term Memory (BiLSTM) model .the system successfully predicts the maximum depth of corrosion based on historical inspection data. The preprocessing phase included rigorous data cleaning, correlation analysis, outlier detection, and temporal feature matching, ensuring the reliability of the training dataset. The predictive model demonstrated strong performance with a low mean absolute error, confirming the suitability of deep learning for handling complex, nonlinear and temporal degradation phenomena.

The integration of such a predictive system offers significant benefits in terms of safety, cost reduction, and real-time maintenance planning. It represents a step forward in the digital transformation of industrial asset monitoring, enabling smarter, data driven decision making. Future improvements may include model optimization, real-time sensor integration, and deployment on low-power embedded hardware for field implementation.

Building on the embedded deployment strategy presented in this chapter, several improvements can be envisioned to enhance the system's performance, intelligence, and scalability. One promising direction is the integration of real-time sensor networks (e.g: temperature, humidity, pressure) directly connected to the Raspberry Pi, enabling the system to continuously monitor environmental and operational conditions that influence corrosion processes. This would enrich the prediction context and allow for more dynamic and responsive corrosion assessment.

In parallel, the deployment of more advanced deep learning models, such as attention-enhanced BiLSTM or hybrid architectures combining convolutional and recurrent layers (CNN-LSTM), could further improve prediction accuracy and generalization across different pipeline configurations.

Finally, extending the current architecture with cloud connectivity and edge-to-cloud collaboration would allow for centralized data aggregation, real-time anomaly reporting, and intelligent maintenance planning. This hybrid approach would preserve the benefits of local

inference while enabling large-scale monitoring and coordination making the system a robust, scalable solution for predictive maintenance in oil and gas pipeline networks.

Bibliographies

Ben Seghier, M. E. A., Höche, D., & Zheludkevich, M. (2022). Prediction of the internal corrosion rate for oil and gas pipeline: Implementation of ensemble learning techniques. *Journal of Natural Gas Science and Engineering*, *99*, 104425. https://doi.org/10.1016/j.jngse.2022.104425

Ben Zhou, & C. Cervantes. (2016, September 8). *Machine Learning Fall-Decision Trees*. https://www.researchgate.net/publication/336348988_Stress_Corrosion_Cracking_Damages

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Canonaco, G., Roveri, M., Alippi, C., Podenzani, F., Bennardo, A., Conti, M., & Mancini, N. (2022). A transfer-learning approach for corrosion prediction in pipeline infrastructures. *Applied Intelligence*, *52*(7), 7622–7637. https://doi.org/10.1007/s10489-021-02771-y

Ch10-la-tenue-a-la-corrosion-de-l-aluminum-en-milieu-marin.pdf. (n.d.). Retrieved March 18, 2025, from https://www.almet-marine.com/wp-content/uploads/2020/05/Ch10-la-tenue-a-la-corrosion-de-l-aluminum-en-milieu-marin.pdf?utm_source=chatgpt.com

Corrosion: Définition, causes, effets et prévention. (n.d.). Camfil. Retrieved March 18, 2025, from https://www.camfil.com/fr-fr/infos/electronique-optique/effets-de-la-corrosion

Du, X., Dai, Y., Zheng, B., & zhu, X. (2024). Corrosion Prediction Model for Long-Distance Pipelines Based on Nonlinear Feature Expansion-Northern Goshawk Optimization-Extreme Learning Machine. *Corrosion*, *80*(11), 1118–1127. https://doi.org/10.5006/4582

Factors influencing Rate of Corrosion—Sealxpert. (2017, May 15). https://www.sealxpert.com/factors-influencing-rate-of-corrosion

Fang, J., Cheng, X., Gai, H., Lin, S., & Lou, H. (2023). Development of machine learning algorithms for predicting internal corrosion of crude oil and natural gas pipelines. *Computers & Chemical Engineering*, *177*, 108358. https://doi.org/10.1016/j.compchemeng.2023.108358

Feng, Y., Siow, K.-S., Teo, W.-K., Tan, K.-L., & Hsieh, A.-K. (1997). Corrosion Mechanisms and Products of Copper in Aqueous Solutions at Various pH Values. *Corrosion*, *53*(5), 389–398. https://doi.org/10.5006/1.3280482

Hamdy Makhlouf, A. S. (2015). Chapter 15—Intelligent Stannate-Based Coatings of Self-Healing Functionality for Magnesium Alloys. In A. Tiwari, J. Rawlins, & L. H. Hihara (Eds.), *Intelligent Coatings for Corrosion Control* (pp. 537–555). Butterworth-Heinemann. https://doi.org/10.1016/B978-0-12-411467-8.00015-5

Ho, T. K. (1995). Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, *1*, 278–282 vol.1. https://doi.org/10.1109/ICDAR.1995.598994

Hossen, A., Mahmud, R., Islam, A., Ahsan, S. K., & Mondal, M. I. H. (2023). Minimization of corrosion in aquatic environment–a review. *Int J Hydro*, *7*(1), 9–16.

Hussein Khalaf, A., Xiao, Y., Xu, N., Wu, B., Li, H., Lin, B., Nie, Z., & Tang, J. (2024). Emerging AI technologies for corrosion monitoring in oil and gas industry: A comprehensive review. *Engineering Failure Analysis*, *155*, 107735. https://doi.org/10.1016/j.engfailanal.2023.107735

Imran, M. M. H., Jamaludin, S., & Mohamad Ayob, A. F. (2024). A critical review of machine learning algorithms in maritime, offshore, and oil & gas corrosion research: A comprehensive analysis of ANN and RF models. *Ocean Engineering*, *295*, 116796. https://doi.org/10.1016/j.oceaneng.2024.116796

Joseph, O. O., Banjo, S., Afolalu, S. A., & Babaremu, K. O. (2022). Global and economic effects of corrosion—An overview. *AIP Conference Proceedings*, *2437*(1), 020149. https://doi.org/10.1063/5.0092286

Kashgain. (n.d.). *The Impact of Corrosion on Infrastructure: Bridges, Pipelines, and Buildings*. Retrieved March 19, 2025, from http://kashgain.net/blog/the-impact-of-corrosion-on-infrastructure-bridges-pipelines-and-buildings/?utm_source=chatgpt.com

Khalifeh, A. (2019). Stress Corrosion Cracking Damages. In Z.-M. Huang & S. Hemeda (Eds.), *Failure Analysis*. IntechOpen. https://doi.org/10.5772/intechopen.80826

Khoshnaw, F., & Gubner, R. (Eds.). (2020). Part I: General Aspects of Corrosion, Corrosion Control, and Corrosion Prevention. In *Corrosion Atlas Case Studies* (pp. xxv–xli). Elsevier. https://doi.org/10.1016/B978-0-12-818760-9.02002-X

Koch, G. H., Brongers, M. P. H., Thompson, N. G., Virmani, Y. P., & Payer, J. H. (2016). *Corrosion Costs and Preventive Strategies in the United States*. https://impact.nace.org/documents/ccsupp.pdf

Lara-Benítez, P., Carranza-García, M., Luna-Romera, J. M., & Riquelme, J. C. (2020). Temporal Convolutional Networks Applied to Energy-Related Time Series Forecasting. *Applied Sciences*, *10*(7), Article 7. https://doi.org/10.3390/app10072322

Lee, L. H., Rajkumar, R., Lo, L. H., Wan, C. H., & Isa, D. (2013). Oil and gas pipeline failure prediction system using long range ultrasonic transducers and Euclidean-Support Vector Machines classification approach. *Expert Systems with Applications*, *40*(6), 1925–1934. https://doi.org/10.1016/j.eswa.2012.10.006

Léopold. (2024, April 9). Tout sur les enjeux de la corrosion en milieu industriel. *Le repère des capitaines d'industrie*. https://www.s20industries.fr/comprendre-corrosion-milieu-industriel/

Li, X., Deng, S., & Lin, C. (2019). *Electrochemical techniques for corrosion monitoring: A review*. *35*(4), 707–720. https://doi.org/10.1016/j.jmst.2018.11.024

Li, X., Guo, M., Zhang, R., & Chen, G. (2022). A data-driven prediction model for maximum pitting corrosion depth of subsea oil pipelines using SSA-LSTM approach. *Ocean Engineering*, *261*, 112062. https://doi.org/10.1016/j.oceaneng.2022.112062

Liu, Y.-W., Zhang, J., Lu, X., Liu, M.-R., & Wang, Z.-Y. (2020). Effect of Metal Cations on Corrosion Behavior and Surface Structure of Carbon Steel in Chloride Ion Atmosphere. *Acta Metallurgica Sinica (English Letters)*, *33*(9), 1302–1310. https://doi.org/10.1007/s40195-020-01032-0

Luo, Z., Hu, X., & Gao, Y. (2013). *Corrosion Research of Wet Natural Gathering and Transportation Pipeline Based on SVM*. 964–972. https://doi.org/10.1061/9780784413142.102

Ma, H., Zhang, W., Wang, Y., Ai, Y., & Zheng, W. (2023). Advances in corrosion growth modeling for oil and gas pipelines: A review. *Process Safety and Environmental Protection*, *171*, 71–86. https://doi.org/10.1016/j.psep.2022.12.054

Madamanchi, A., Rabbi, F., Sokolov, A. M., & Hossain, N. U. I. (2024a). A Machine Learning-Based Corrosion Level Prediction in the Oil and Gas Industry. *Engineering Proceedings*, *76*(1), Article 1. https://doi.org/10.3390/engproc2024076038

Madamanchi, A., Rabbi, F., Sokolov, A. M., & Hossain, N. U. I. (2024b). A Machine Learning-Based Corrosion Level Prediction in the Oil and Gas Industry. *ICIMP 2024*, 38. https://doi.org/10.3390/engproc2024076038

Malashin, I., Tynchenko, V., Nelyub, V., Borodulin, A., Gantimurov, A., Krysko, N. V., Shchipakov, N. A., Kozlov, D. M., Kusyy, A. G., Martysyuk, D., & Galinovsky, A. (2024). Deep Learning Approach for Pitting Corrosion Detection in Gas Pipelines. *Sensors (Basel, Switzerland)*, *24*(11), 3563. https://doi.org/10.3390/s24113563

Mémoire de Fin d études. (n.d.). Retrieved March 18, 2025, from https://123dok.net/document/yngd2mmp-m%C3%A9moire-de-fin-d-%C3%A9tudes.html

Mysliu, E., Taji, I., & Erbe, A. (2025). A review of filiform corrosion and its prevention on polymer-coated aluminium alloys. *RSC Applied Interfaces*, *2*(2), 304–319. https://doi.org/10.1039/D4LF00231H

Poursaee, A. (2023). 15—Corrosion protection methods of steel in concrete. In A. Poursaee (Ed.), *Corrosion of Steel in Concrete Structures (Second Edition)* (pp. 319–326). Woodhead Publishing. https://doi.org/10.1016/B978-0-12-821840-2.00014-6

Prasetia, Y. (2025, January 10). Corrosion Mechanisms Understanding Causes and Effects. *PetroSync Blog*. https://www.petrosync.com/blog/corrosion-mechanisms/

Raja, V. S., & Shoji, T. (2021). *Stress Corrosion Cracking: Theory and Practice* (Elsevier). https://www.elsevier.com/books/stress-corrosion-cracking/raja/978-0-08-101641-9

Rao, P., & Mulky, L. (2023). Microbially Influenced Corrosion and its Control Measures: A Critical Review. *Journal of Bio- and Tribo-Corrosion*, *9*(3), 57. https://doi.org/10.1007/s40735-023-00772-7

Ravi Kumar Gali & Ramesh Kumar V. (2023). Corrosion studies of various salt solutions on metals and alloys. *Global Journal of Engineering and Technology Advances*, *16*(3), 007–018. https://doi.org/10.30574/gjeta.2023.16.3.0174

Sheryazov, S. K., Sarkulova, Z. S., Balgynova, A. M., Shukirova, S. S., & A.zh, T. (2024). METHODS OF CORROSION PROTECTION OF EQUIPMENT AND PIPELINES IN THE OIL AND GAS INDUSTRY. *Научный Журнал "Вестник Актюбинского Регионального Университета Имени К. Жубанова,"76*(2), Article 2. https://doi.org/10.70239/arsu.2024.t76.n2.05

Son, E.-Y., Jeong, D., & Oh, M.-J. (2024). Corrosion area detection and depth prediction using machine learning. *International Journal of Naval Architecture and Ocean Engineering*, *16*, 100617. https://doi.org/10.1016/j.ijnaoe.2024.100617

Soomro, A. A., Mokhtar, A. A., Kurnia, J. C., Lashari, N., Lu, H., & Sambo, C. (2022a). Integrity assessment of corroded oil and gas pipelines using machine learning: A systematic review. *Engineering Failure Analysis*, *131*, 105810. https://doi.org/10.1016/j.engfailanal.2021.105810

Soomro, A. A., Mokhtar, A. A., Kurnia, J. C., Lashari, N., Lu, H., & Sambo, C. (2022b). Integrity assessment of corroded oil and gas pipelines using machine learning: A systematic review. *Engineering Failure Analysis*, *131*, 105810. https://doi.org/10.1016/j.engfailanal.2021.105810

Sooriyalakshmi, N., & Helena, H. J. (2021). *Study on Rebar Corrosion Monitoring and Prevention Techniques*. https://www.academia.edu/download/88343049/Sooriyalakshmi_2022_BP_2185A.pdf

Sylvain.S.39. (2024, October 15). Pollution des sols: Des répercussions cachées sur notre écosystème. *Nature et Faune*. https://natureetfaune.com/menaces-environnementales/les-consequences-de-la-pollution-des-sols-sur-lenvironnement/

Tait, W. S. (2018). Chapter 27—Controlling Corrosion of Chemical Processing Equipment. In M. Kutz (Ed.), *Handbook of Environmental Degradation of Materials (Third Edition)* (pp. 583–600). William Andrew Publishing. https://doi.org/10.1016/B978-0-323-52472-8.00028-9

Wang, C., Ma, G., Li, J., Dai, Z., & Liu, J. (2019). Prediction of Corrosion Rate of Submarine Oil and Gas Pipelines Based on IA-SVM Model. *IOP Conference Series: Earth and Environmental Science*, *242*, 022023. https://doi.org/10.1088/1755-1315/242/2/022023

Wang, H. (2023). Corrosion Theory and Corrosion Characterization Techniques. In L. Zhang (Ed.), *Corrosion in CO2 Capture, Transportation, Geological Utilization and Storage: Causes and Mitigation Strategies* (pp. 9–30). Springer Nature. https://doi.org/10.1007/978-981-99-2392-2_2

Wang, J., Kong, Z., Shan, J., Du, C., & Wang, C. (2024a). Corrosion Rate Prediction of Buried Oil and Gas Pipelines: A New Deep Learning Method Based on RF and IBWO-Optimized BiLSTM–GRU Combined Model. *Energies*, *17*(23), 5824. https://doi.org/10.3390/en17235824

Wang, J., Kong, Z., Shan, J., Du, C., & Wang, C. (2024b). Corrosion Rate Prediction of Buried Oil and Gas Pipelines: A New Deep Learning Method Based on RF and IBWO-Optimized BiLSTM–GRU Combined Model. *Energies*, *17*(23), 5824. https://doi.org/10.3390/en17235824

Wang, Q., Song, Y., Zhang, X., Dong, L., Xi, Y., Zeng, D., Liu, Q., Zhang, H., Zhang, Z., Yan, R., & Luo, H. (2023). Evolution of corrosion prediction models for oil and gas pipelines: From empirical-driven to data-driven. *Engineering Failure Analysis*, *146*, 107097. https://doi.org/10.1016/j.engfailanal.2023.107097

Wang, Z., Wang, Q., Jia, C., & Bai, J. (2022). Structure characteristics and evolution mechanism of oil sands bitumen at Karamay, Xinjiang(China). *Journal of Petroleum Science and Engineering*, *214*, 110421. https://doi.org/10.1016/j.petrol.2022.110421

Wasim, M., Shoaib, S., Mubarak, N. M., Inamuddin, & Asiri, A. M. (2018). Factors influencing corrosion of metal pipes in soils. *Environmental Chemistry Letters*, *16*(3), 861–879. https://doi.org/10.1007/s10311-018-0731-x

Xu, L., Wang, Y., Mo, L., Tang, Y., Wang, F., & Li, C. (2023a). The research progress and prospect of data mining methods on corrosion prediction of oil and gas pipelines. *Engineering Failure Analysis*, *144*, 106951. https://doi.org/10.1016/j.engfailanal.2022.106951

Xu, L., Wang, Y., Mo, L., Tang, Y., Wang, F., & Li, C. (2023b). The research progress and prospect of data mining methods on corrosion prediction of oil and gas pipelines. *Engineering Failure Analysis*, *144*, 106951. https://doi.org/10.1016/j.engfailanal.2022.106951

Xu, W., Zhang, R., Duan, J., & Greenfield, D. T. (2024). Corrosion is a global menace to crucial infrastructure—Act to stop the rot now. *Nature*, *629*(8010), 41–41. https://doi.org/10.1038/d41586-024-01270-7

Yarveisy, R., Khan, F., & Abbassi, R. (2024). Data-driven model to predict burst pressure in the presence of interacting corrosion pits. *Journal of Pipeline Science and Engineering*, *4*(1), 100146. https://doi.org/10.1016/j.jpse.2023.100146

Zadeh, S. A. (2023). Erosion-corrosion. In *Introduction to Corrosion—Basics and Advances*. IntechOpen. https://doi.org/10.5772/intechopen.109106

Zehra, S., Mobin, M., & Aslam, R. (2022). CHAPTER 2—Corrosion prevention and protection methods. In L. Guo, C. Verma, & D. Zhang (Eds.), *Eco-Friendly Corrosion Inhibitors* (pp. 13–26). Elsevier. https://doi.org/10.1016/B978-0-323-91176-4.00023-4

Dann, M. R., & Dann, C. (2017). Automated matching of pipeline corrosion features from in-line inspection data. *Reliability Engineering & System Safety*, *162*, 40–50. https://doi.org/10.1016/j.ress.2017.01.008

Dinesh Naik, & C. D. Jaidhar. (2022). (PDF) A novel Multi-Layer Attention Framework for visual description prediction using bidirectional LSTM. *ResearchGate*. https://doi.org/10.1186/s40537-022-00664-6

Yarveisy, R., Khan, F., & Abbassi, R. (2021). *Dataset for: Cross-country Pipeline Inspection Data Analysis and Testing of Probabilistic Degradation Models*. *1*. https://doi.org/10.17632/c2h2jf5c54.1